

Tallinna Pedagoogikaülikool
Infoteaduste osakond

RAAMATUKOGU
ELEKTRONKATALOOG JA AINEOTSING:
ARENDUSPROBLEEME 1990. AASTATE EESTIS
Magistritöö

SIRJE NILBE

Juhendaja: prof. emer Evi Rannap

Tallinn 2004

SISUKORD

SISSEJUHATUS	5
1. RAAMATUKOGU ELEKTRONKATALOOG	9
1.1. Elektronkataloogide areng	9
1.2. Elektronkataloogide juurutamine Eestis	13
1.3. Elektronkataloogi kontseptuaalne mudel	14
1.4. Elektronkataloogi andmebaas	15
2. INDEKSEERIMISKEELED JA AINEOTSING	20
2.1. Liigitamine ja märksõnastamine aineotsingu alusena	20
2.2. Märksõnastamine elektronkataloogis	21
2.2.1. Märksõnastamise meetodid	21
2.2.2. Tesauruse olemus ja ülesehitus	23
2.2.3. Tesaurus ja arvuti	26
2.3. Universaalse kümnendliigituse (UDK) rahvusvahelisest levikust	28
2.3.1. UDK rakendusala	28
2.3.2. UDK levik maailma maades	31
2.4. Indekseerimise ja indekseerimiskeelte standardimine	32
3. KOKKUVÕTE JA JÄRELDUSED	34
KASUTATUD KIRJANDUS	37
ONLINE CATALOGUES AND SUBJECT ACCESS: DEVELOPMENT	
PROBLEMS IN ESTONIA IN 1990s	39
PUBLIKATSIOONID	42

PUBLIKATSIOONIDE LOETELU

Magistritöö põhineb järgmistel artiklitel.

1. Nilbe, S. Raamatukogu elektronkataloog 1. arengulugu ja olemus. *Infofoorum*, 2, veebr. 1997
<http://www.tpu.ee/~i-foorum/> (18.05.04)
2. Nilbe, S. Raamatukogu elektronkataloog 2: andmebaas. *Infofoorum*, 4, juuli 1998.
<http://www.tpu.ee/~i-foorum/> (18.05.04)
3. Loorits, E., Nilbe, S. Kataloogimise automatiseerimisest Tartu Ülikooli raamatukogus. *Raamatukogu*, 1995(1), 9-11.
4. Nilbe, S. Tesaauruse mitu palet. *Eesti Raamatukoguhoidjate Ühingu aastaraamat 7 (1995)*. Tallinn, 1996, lk. 51-67
5. Nilbe, S. Thesaurus for subject searching and indexing as part of an integrated library system. *International conference on library automation in Central and Eastern Europe: conference proceedings, Budapest 10-13 April 1996*. Luxembourg, 1997, pp. 157-159.
6. Nilbe, S. Märksõnastamine Tartu Ülikooli raamatukogus. *Raamatukogu*, 1997(2), 38-39.
7. Nilbe, S. Universaalse kümnendliigituse (UDK) rahvusvahelisest levikust. *Tartu Ülikooli raamatukogu töid 10*. Tartu, 1997, lk. 169-183.
8. Nilbe, S. Standardization in subject indexing: possibilities and activities. *Independence and libraries: papers of the 5th Congress of Baltic Librarians, October 21-22 1996*. Tallinn, 1996, pp. 107-111.
9. Nilbe, S. Dokumentide sisu avamine ja aineotsing arvutikataloogis. *Raamatukogu*, 1995(1), 12-13.

Publikatsioonide kogumaht on 62 lk.

Autori avaldamata ettekanded, mis on seotud magistritöö temaga.

1. Tietokantaan integroitu hakutesaurus kirjaston näyttöluettelossa. Oulu Ülikooli infoteaduse ja sotsioloogia osakonna seminaril 18.05.1995.
2. Teavikute sisuanalüüs ja aineotsing õppeainena. Tartu Ülikooli raamatukogu seminaril „Raamatuteaduslik ja raamatukogunduslik kõrgharidus” 06.12.1996.
3. Autoriteetfailide hooldamine INNOPAC-süsteemis. Teaduslike keskraamatukogude suveseminaril „Integreeritud raamatukogusüsteem INNOPAC ja teadusraamatukogude koostöö” 18.08.1997
4. Cataloguing: current state and future perspectives in Estonia. Baltic Practical Scientific Seminar on Cataloguing, September 8-12, 1997, Vilnius. (Teesid avaldatud.)
5. Elektronkataloogilgi on oma ajalugu: INGRIDist INNOPACini. Tartu Ülikooli raamatukogu XVII teaduskonverentsil 24.09.1997
6. INNOPACi juurutamise seis: normandmed ja otsinguvõimalused. Teadusraamatukogude suveseminaril 10.08.1998.
- 7 Töö normandmetega Rahvusraamatukogus. Teaduskonverentsi „Rahvus ja raamatukogu” sektsioonis „Informatsioonilise infrastruktuuri areng Rahvusraamatukogus” 18.12.1998. (Teesid avaldatud.)

SISSEJUHATUS

Käesolev magistritöö annab ülevaate autori uurimis- ja arendustegevusest elektronkataloogide juurutamise ajal Eestis 1990. aastatel. Minu põhitähelepanu oli suunatud elektronkataloogis kasutatavatele indekseerimiskeeltele, sisukirjelduse meetoditele ja aineotsingu probleemidele.

Indekseerimiskeele all mõistan dokumendi sisu või vormi iseloomustamiseks loodud tehiskeelt. Sellise määratluse annab terminile standard ISO:5127 „Information and documentation. Vocabulary”, mille Eesti vaste EVS-ISO 5127 on kohe valmimas. Indekseerimiskeeli kasutatakse kirjeldamiseks, mille *kohta* dokument on. Selle tegevuse eesmärgiks on sarnase infosisuga dokumentide koondamine. Teiseks võimaldavad indekseerimiskeeled kasutajal navigeerida bibliograafilises *resp* teadmiste universumis (Svenonius 2000, 127).

Sisukirjeldus on ISO:5127 järgi indekseerimiskeelde kuuluvatest märksõnadest või notatsioonisümbolitest koosnev kirjeldus, mis põhineb dokumendis sisalduvatel andmetel. Tavaliselt peetakse sisukirjelduseks siiski ka sisu tekstilisi esitusi, nt annotatsioone ja referaate.

Aineotsing e teemaotsing on infootsingu liik, mille puhul infovajaduse rahuldamiseks otsitakse seni teadmata infoallikaid (dokumente) nende sisuliste, ainestikust lähtuvate kriteeriumide alusel (Hjørland 1997, 21).

Elektronkataloogide aineotsingu võimalusi kritiseeriti 1980. aastate teisel poolel ja 1990. aastate alguses erialakirjanduses sageli. Larsoni järgi (1991, 178-190) saab kasutajate probleemid kokku võtta järgmiselt.

1. Lugejad ei tunne sisukirjelduseks kasutatud märksõnastikke ega liigitussüsteeme.
2. Lugejad ei oska päringut õigesti moodustada. Muuhulgas ei saa nad aru Boole'i loogikast ja teevad rohkesti kirjavigu.
3. Lugejad ei oska formuleerida uut päringut, kui esimene otsing on toonud kaasa liiga suure või liiga napi kirjade arvu.

4. Ka mahult paras kirjete arv võib olla lugejale kasutu või väheväärtuslik, kui ta ei ole leidnud sobivat otsiterminite.

Näib nii, et aineotsingu suurim probleem on saada vastavusse dokumendi, indekseerija kui vahendaja ja infootsija kasutatav sõnavara. Selle saavutamiseks on tehtud mitmesuguseid ettepanekuid (Larson 1991; Norgard jt 1993; O'Brien 1994; Su 1994):

- 1) lisada bibliokirjetele rohkem märksõnu ja liigiindekseid
- 2) täiendada kirjeid sõnadega teavikute sisukordadest ja registritest
- 3) täiendada kataloogi registreid kasutajate otsisõnadega
- 4) lisada ohjatud märksõnaregistritesse võtmesõnu nt pealkirjadest
- 5) otsisõnade automaatne kärpimine või tüvestamine
- 6) automaatsed ristviited
- 7) päringute õigekirjakontroll
- 8) teasauruste, liigitustabelite ja sõnastike ühendamine andmebaasiga
- 9) liigiindeksite ja märksõnade ühendamine (klasteriseerimine)
- 10) eri märksõnastike ühendamine

Need ettepanekud tehti parandamiseks aineotsingut juba olemas olevates suure andmebaasiga elektronkataloogides, kus indekseerimissüsteemi kardinaalne muutmine ei olnud võimalik. Eesti raamatukogud olid elektronkataloogimist alles alustamas ja oli võimalik oma tegevust paremini planeerida.

Seega oli magistr töö teema ajal, mil hakkasin sellega tegelema, ka Eestis väga aktuaalne. Magistr töö osana esitatud artiklid olid Eestis esimesed, milles käsitleti aineotsingu kujundamist elektronkataloogides. Hiljem on teema mõningaid aspekte analüüsinud Kruus (2000), käsitledes peamiselt tesauruse rolli infootsingus ja UDKd infootsingu vahendina andmebaasides.

Magistr töö eesmärk oli leida Eesti raamatukogudele sobivad praktilised mudelid sisukirjelduse ja aineotsingu jaoks elektronkataloogides, lähtudes rahvusvahelisest kogemusest ja meie enda ideedest.

Eesmärgi täitmiseks tuli lahendada rida ülesandeid.

- 1) uurida elektronkataloogi olemust integreeritud raamatukogusüsteemi põhikomponendi ja infootsüsteemina
- 2) analüüsida standardse bibliokirje elemente ja vormingut aineotsingu võimaluste seisukohalt
- 3) uurida elektroonilisse keskkonda sobivate märksõnastike tüüpe ning märksõnastamise metoodikat
- 4) selgitada universaalse detsimaalklassifikatsiooni seisundit maailmas ja selle jätkusuutlikkust
- 5) leida indekseerimist puudutavad rahvusvahelised standardid ja analüüsida nende rakendatavust
- 6) uurida aineotsingu kujundamise võimalusi ja hinnata eri otsimeetodite otstarbekat vahekorda.

Uurimisobjektiks said elektronkataloogid ja indekseerimiskeeled ning uurimisaineks indekseerimiskeelte funktsioneerimine aineotsingus elektroonilises keskkonnas. Uurimismeetoditena kasutasin teemakohase erialakirjanduse ja tehnilise dokumentatsiooni analüüsi, indekseerimiskeelte struktuuri ja indekseerimismeetodite võrdlevat analüüsi ning reaalse infootsüsteemide toimimise analüüsi.

Kuna uurimistöö eesmärk on vahetult seotud praktilise raamatukogundusega, olen selle tulemusi mitmel moel rakendatud: märksõnastike koostamisel Tartu Ülikooli raamatukogus ja Rahvusraamatukogus, eestikeelse UDK väljaande koostamisel, aineotsingu võimaluste kujundamisel Tartu Ülikooli elektronkataloogis INGRID, ELNET Konsortsiumi infootsüsteemis ESTER, rahvaraamatukogude infootsüsteemides jm. Kujunenud seisukohti ja avaldatud publikatsioone on kasutatud nii autori enda kui ka teiste õppejõudude ja koolitajate läbiviidud õppetöös.

Olen õpetanud liigitamist ja indekseerimist Tallinna Pedagoogikaülikooli infoteaduste osakonna bakalaureuseõppes, raamatukoguhoidjate kutseõppe kursustel ning arvukatel muudel koolitusüritustel. Aastast 1993 olen Eesti Raamatukoguhoidjate Ühingu liigitamise ja märksõnastamise toimkonna liige, aastast 1997 ELNET Konsortsiumi liigitamise ja märksõnastamise töörühma juht ning tehnilise töörühma liige.

Magistritöö koosneb kahest osast: “Raamatukogu elektronkataloog” ning “Indekseerimiskeeled ja aineotsing”. Analüütilisele ülevaatele on lisatud publikatsioonid, milles on avaldatud enamik uurimistöö tulemusi.

1 RAAMATUKOGU ELEKTRONKATALOOG

1.1 Elektronkataloogide areng

Elektronkataloogide ja teiste sidusandmebaaside juurutamine Eesti raamatukogudes algas 1990. aastate esimesel poolel. Esimesed andmebaasid olid mõeldud eelkõige bibliograafiaandmete säilitamiseks trükibibliograafiate ettevalmistamise või infoteeninduse eesmärgil, mitte lugejatele infootsinguks. Selles osas ei erinenud Eesti arengud muust maailmast, ainult et meil algas infotehnoloogia rakendamine raamatukogudes hiljem ja kulges kiiremini.

Lugeja infootsingu seisukohalt vaadates algas uus etapp raamatukogude automatiseerimises avalike elektronkataloogide evitamisega.

Raamatukogude elektronkatalooge saab vaadelda kahest aspektist: kui infootsisüsteeme ja kui raamatukogutöö automatiseerimise põhikomponenti. Infootsisüsteemina peab elektronkataloog täitma eelkõige neidsamu nõudeid, mis C. A. Cutter juba 1867. aastal raamatukogukataloogile esitas. Kataloog peab

- 1) võimaldama leida raamatut, millest on teada kas autor, pealkiri või teema
- 2) näitama, mis teoseid raamatukogus leidub teatud autorilt, teatud teemal või teatud laadi kirjandusest
- 3) abistama teose valikul kas bibliograafiliste, temaatiliste või kirjanduslike tunnuste järgi.

Raamatukogutöö tehnoloogia seisukohalt on elektronkataloogi andmebaas selleks tuumaks, millel rajaneb kogude haldus ja laenutus. Elektronkataloogide sünnis oli määrava tähtsusega just viimane aspekt – vajadus ratsionaliseerida kataloogimist ja laenutuste arvestust.

Kui kataloogima hakati arvuti abil, ei nähtud algul ette tekkiva andmebaasi kasutamist lugeja poolt. See oli mõeldud raamatukogu sisetööks, et hõlbustada traditsiooniliste kataloogivormide – kaart-, trüki- ja mikrovormkataloogide tootmist. Üsna varsti leiti, et need andmebaasid võiksid olla kättesaadavad ka lugejale nii raamatukogus kui

sideliinide kaudu väljaspool raamatukogu. Esimesed automatiseeritud süsteemid, mida võiks nimetada elektronkataloogideks, võeti kasutusele 1970. aastate keskel Ameerika Ühendriikides.

Praktilisest seisukohast on kõige olulisemad erinevused kaart- ja elektronkataloogi vahel järgmised.

1. Elektronkataloogis on koos raamatukogu kõik varad, lugeja ei pea eri teavikulaade otsima eri kataloogidest.
2. Elektronkataloog on kättesaadav paljudes kohtades raamatukogus ja ka väljaspool raamatukogu infovõrkude kaudu.
3. Elektronkataloogi kirjetes on rohkem otsitunnuseid kui kaartkataloogis.
4. Elektronkataloogi abil saab lugeja uued materjalid kiiremini kätte, sest kirjed liituvad andmebaasi vahetult pärast nende koostamist.
5. Elektronkataloogis on iga teaviku kohta vaja koostada vaid üks kirje.
6. Elektronkataloogi haldamine (kirjete liitmine, kustutamine, parandamine, väljavõtete tegemine) on jõudsam kui kaartkataloogi puhul.
- 7 Elektronkataloogi saab siduda teiste andmebaaside ja infosüsteemidega, tekitades inforuumi, mis ulatub kaugele väljapoole raamatukogu piire.

C. R. Hildreth, juhtivaid uurijaid sel alal, on juba 1983. aastast alates korduvalt rääkinud ja kirjutanud elektronkataloogide kolmest põlvkonnast. Tema jaotus põhineb kataloogide kvalitatiivsetel erinevustel, sidumata ühtegi põlvkonda mingi kindla ajavahemikuga. Lühidalt võib Hildrethi käsitlese elektronkataloogidest kokku võtta järgmiselt (Hildreth 1987).

Esimeste elektronkataloogide kirjed olid üsna napid: tihti vaid autor, pealkiri, kohaviit ja saadavusandmed. Märksõnu sageli polnudki. Ei olnud võimalik kasutada vabatekstiotsingut ega kärpimist, päring pidi täpselt vastama andmebaasi registri sõnadele või väljenditele. Järelkoordinatsiooni ei saanud üldiselt rakendada, mõnes süsteemis siiski võis otsida koos autori nime ja pealkirja algust. Esimese põlvkonna elektronkataloogid eeldasid, et kasutaja teab täpselt, mida otsib. Registrite või kirjete sirvimise võimalust polnud või oli see väga piiratud. Otsiterminate valimiseks ega muukski süsteem abi ei pakkunud ja tihtipeale lugeja ei saanudki aru, miks tema

otsing ebaõnnestus. Suhtlemine arvutiga käis enamasti menüürežiimis ja tulemuseks saadi lühikirjed, mis võimaldasid teose lokaliseerida.

Teise põlvkonna elektronkataloogides on esimese põlve vigu püütud parandada, võttes eeskujuks infootsisüsteemide omadusi. Kirjed on täielikumad ja otsitavaid välju on rohkem. On suurendatud registrite arvu ja täiustatud normandmefailide ning antud võimalus neid sirvida. Aineotsing märksõnade ja võtmesõnade abil on tavaline ning täiendatud järelkoordinatsiooni ning Boole'i operaatoritega. Võtmesõnaregister on moodustatud tavaliselt pealkirja ja märksõnade väljadest. Otsitermineid saab paremalt poolt kärpida. Sageli on võimalik otsingut piirata ilmunisaja, ilmuniskoha või keele järgi. Kasutajaliides on muutunud mitmekesisemaks ja interaktiivsemaks: juba registris antakse teada otsitunnusele vastavate kirjete arv, otsingu käigus on võimalik tagasi liikuda, saab valida menüürežiimi (algajatele) ja käsürežiimi (edasijõudnutele) vahel jms. Leitud kirjed on võimalik saada mitmes vormis (nt lühikirje, täiskirje, MARC-kirje). Niisiis on teise põlvkonna elektronkataloogides kui infootsisüsteemides ühendatud hästistruktureeritud raamatukogukataloogi ja tüüpilise Boole'i otsingul põhineva infootsisüsteemi omadused.

Hildrethi arvates peaksid elektronkataloogid muutuma veel efektiivsemaks ja kasutajasõbralikumaks, kui lisada teise põlve omadustele mitmeid täiendusi, ning moodustama siis kolmanda põlvkonna. Ta peab oluliseks, et päringuid saaks esitada loomulikus keeles, et andmebaasis navigeerimist assisteeriks intelligentsed programmid, et otsingu igas järgus oleks hea tagasiside. Aineotsingus näiteks tuleks integreerida märksõna-, liigi- ja võtmesõnaotsing, nii et otsiterminite konversioon toimiks automaatselt. "Täpse vaste" (*exact-match*) printsiibi asemel peaks süsteem toimima "lähima vaste" (*best-match*) põhimõtte järgi. Väljundkirjed tuleks reastada vastavalt nende oletatavale relevantsusele ja kasutaja peaks saama anda tagasisidet väljundi relevantsuse kohta.

Kataloog peaks hõlmama raamatukogu varasid täielikult, osalt bibliokirjetena, osalt täistekstide ja kujutistena. Bibliokirjed võiksid olla inforikkamad. Elektronkataloogile peaks olema laialdane juurdepääs infovõrkude kaudu, teisalt peaks kataloog olema seotud mitmesuguste teiste andmebaaside ja infosüsteemidega.

Kolmanda põlvkonna elektronkatalooge, millel oleks kasvõi enamik loetletud omadusi, ei eksisteeri siia maani. Prototüüpsüsteeme, milles katsetatakse üht või mitut uuendust, on küll loodud, kuid raamatukogusüsteeme tootvad firmad on üsna visad oma toodangut kardinaalselt muutma. Integreeritud raamatukogusüsteem on keerukas infosüsteem, ja raamatukogud mitte just eriti maksujõulised kliendid.

Raamatukogu seisukohalt vaadates leiab samuti mitmeid takistusi elektronkataloogide põhimõtteliseks uuendamiseks. Kataloog on kumulatiivne andmebaas, mille kirjete, sisukirjelduse ja normandmefailide tagantjärele täiendamine ei tule suuremas mahus kõne alla, eriti kui järjekorras ootamas hulk konverteerimata materjali ja uudiskirjandust. Kui on vaja valida, kas eelistada sisetöö ja bibliograafilise kontrolli või lugeja vajadusi, kaldub kogu maailma raamatukoguhoidjate südametunnistus esimese alternatiivi poole, kasvõi juba seepärast, et oma vajadused on paremini ja ühemõttelisemalt teada.

Veel aastaid pärast eelpool viidatud artikli ilmunist toob Hildreth välja paljusi needsamad vajakajäämised elektronkataloogide funktsionaalsuses (Hildreth 1995): võimatus esitada päringuid loomulikus keeles, süsteemi suutmatus päringut automaatselt teisendada, pole rakendatud "lähima vaste" põhimõtet, puudub väljundi reastamine relevantsuse järgi ning relevantsuse tagasiside, pole rakendatud hüperteksti võimalusi seotud kirjete leidmiseks, võtmesõna-, märksõna- ja liigotsing pole integreeritud, kataloogi ulatus ja haare peaks olema suurem.

Ei saa siiski väita, et elektronkataloogide areng oleks seiskunud. Hüperteksti võimalusi on rakendatud tänapäeval vist küll kõigis veebikataloogides, k.a INNOPACil põhinev Eesti raamatukogude koondkataloog ESTER. INNOPAC/Millennium areneb pidevalt ka selles suunas, et hõlmata ja pakkuda juurdepääsu sellistele inforessurssidele, mis traditsiooniliselt pole kuulunud raamatukogukataloogi. Aga selles on Hildrethil küll õigus, et elektronkataloogides pole siia maani rakendatud uuemaid infootsitehnikaid.

1.2. Elektronkataloogide juurutamine Eestis

Olen 1990. aastatel kaasa aidanud kahe integreeritud raamatukogusüsteemi planeerimisele ja juurutamisele, keskendudes eelkõige kataloogi moodustamise ja kataloogiotsingu, eriti aineotsingu probleemidele.

1993. aastal alustati Tartu Ülikooli raamatukogus esimesena Eestis ja oma jõududega automatiseeritud raamatukogusüsteemi INGRID loomist. See sai tehniliste lahenduste poolest paljuskki ebatüüpiline, eriti angloameerika malliga võrreldes. Süsteemi arendajad toetusid rohkem oma ideedele kui senistele kogemustele mujal maailmas. Ühelt poolt oli see hea, soodustades originaalsete lahenduste loomist (näiteks tesauruse moodul), teiselt poolt tekkis kogenematuse tõttu vigu ja puudujääke, mida põhjalikuma ettevalmistusega oleks saanud vältida. 1994. aasta novembris alustati raamatukogus kataloogimist arvuti abil. INGRIDi kataloogimisüsteemi, selle esimesi kasutuskogemusi ja tulevikuplaane tutvustasin ühisartiklis Elsa Looritsaga (3. artikkel) ajakirjas “Raamatukogu”

INGRIDist sai esimene avalik elektronkataloog ja veebikataloog Eestis.

Paari aasta pärast jõudsid teadusraamatukogud veendumusele, et infotehnoloogilise arengu kiirendamiseks tuleb ühineda ja muretseda läbiproovitud töökindel kommertssüsteem, milleks valiti USA päritolu INNOPAC. INNOPAC sobis oma arhitektuurilt konsortsiumile kõige paremini, võimaldades pidada kahte andmebaasi (Tallinnas ja Tartus), kahte ristikasutusega lugejakataloogi ning vajalikke mooduleid igale raamatukogule eraldi.

Ka Tartu Ülikooli raamatukogu sai Eesti Raamatukoguvõrgu Konsortsiumi liikmeks ning lõpetas INGRIDi arendustööd. 1998. aasta lõpuks, mil alustati kataloogimist INNOPACis, oli INGRID töötanud neli aastat. Kuigi süsteem ei saanudki päris valmis ja sellest tuli lõpuks loobuda, peab tõdema, et need neli aastat ei olnud asjata kulutatud. See oli hindamatu kogemuste omandamise aeg nii raamatukoguhoidjatele, arvutispetsialistidele kui ka lugejatele elektronkataloogiga harjumiseks. Ilma ülikooli

raamatukogu eelneva kogemusega oleks INNOPACi evitamine palju raskemini kulgenud. Tähtsusetu pole ka see, et ülikooli raamatukogu sai kohe uude süsteemi laadida umbes 50 000 korralikku ja sobivalt vormindatud bibliokirjet koos eksemplariandmetega.

INNOPACi evitamine tõstatas vajaduse paremini tundma õppida elektronkataloogide olemust ja funktsioneerimispõhimõtteid, MARC-vormingut ja muid tehnilisi aspekte, et seadistada süsteem meile sobivalt ning kooskõlastada konsortsiumi liikmete vajadusi. Sel perioodil kirjutasin kaks elektronkataloogi-teemalist artiklit elektroonilisele ajakirjale "Infofoorum" (1. ja 2. artikkel). Neis käsitletakse elektronkataloogi kontseptuaalset mudelit ja üksikasjalikumalt andmebaasikihti, kuna kvaliteetse andmebaasi moodustamine on tõhusa ja kasutajat rahuldava kataloogi vältimatu eeldus.

1.3. Elektronkataloogi kontseptuaalne mudel

R. R. Larsoni järgi (Larson 1991) saab iga elektronkataloogi, hoolimata selle konkreetsetest iseärasustest, vaadelda kui funktsionaalsete kihtide süsteemi, mis vahendab kasutajat ja andmebaasi säilitatavaid bibliograafilisi andmeid.

Kasutajaliidese kiht võimaldab lugejal suhelda elektronkataloogi ülejäänud osadega. Kasutajaliidese kihti kuuluvad need süsteemi osad, mis vahetult tegelevad infovahetusega kasutaja ja süsteemi vahel. See on kataloogi kõige nähtavam osa ja mõjutab, nagu näitavad uuringud, kõige enam tavakasutaja arvamust kataloogist, kuigi andmebaasi sisu ja süsteemi otsiomadused on tulemuste seisukohalt olulisemad. Kasutajaliidese põhiülesanded on kasutaja käskude vastuvõtt ja analüüs, selle põhjal süsteemi teiste osade aktiveerimine, väljaotsitud andmete korrastamine, vormindamine ja kuvamine. Hästi projekteeritud ja kergesti käideldav kasutajaliides võib varjata otsimehhanismide sisemise keerukuse või korvata nende mehhanismide puudujääke inimese ja arvuti interaktsiooni parandamise teel.

Andmebaasihaldussüsteemi liidese kiht sisaldab protseduure kasutajaliidese ja andmebaasihaldussüsteemi vaheliseks suhtlemiseks. Selle kihi ülesanne on tõlkida

kasutaja käsud või menüüvalikud vastavaks reaks andmebaasioperatsioonideks. Liidese kihi ülesandeks on ka sellised otsinguga seotud toimingud, mida ei soorita andmebaasihaldussüsteem ise, näiteks loogikaoperatsioonid eelnevalt väljaotsitud kirjete hulkadega või kirjete reastamine relevantsusjärjestuses.

Andmebaasihaldussüsteemi kiht tegeleb andmebaasi salvestatud informatsiooni korrastamise ja taasleidmisega. Selle kihi ülesanneteks on uute kirjete lisamine andmebaasi, kirjete kustutamine ja parandamine ning vastavate muutuste tegemine registritesse, kirjete otsimine ja leidmine otsitunnustele antud väärtuste järgi. Lugejad saavad kasutada vaid elektronkataloogi otsifunktsiooni. Kirjete lisamine, parandamine ja kustutamine on võimalik ainult selleks volitatud personaliliikmetel. Andmebaasi hooldus toimub sõltuvalt tarkvarast kas reaalajas, mispuhul muudatused andmebaasis ilmnevad vahetult pärast nende tegemist, või pakktöötlusena, mil sisestatud muudatused lisatakse andmebaasi kogumina näiteks öhtul või öösel.

Andmebaasikiht säilitab bibliograafilist ja abiinformatsiooni, milleks on tavaliselt bibliokirjed, registrid ja normandmefailid.

Funktsionaalsed kihid on realiseeritud infootsisüsteemina vastava tarkvara ja riistvara abil. Realisatsiooni tasandil on kihtide ülesanded lahendatud eri süsteemides erinevalt ja kihid ise pole nii selgelt eristatavad.

Raamatukogukataloogile omast informatsiooni sisaldab ja säilitab niisiis ainult andmebaasikiht, kuna ülejäänud kihtide ülesanne on seda informatsiooni korrastada ja kasutajale väljastada. Funktsionaalsete kihtide, eriti andmebaasi suhtelise iseseisvuse tõttu on võimalik näiteks kirjeid ühest süsteemist teise üle kanda.

1.4. Elektronkataloogi andmebaas

Kõige tähtsamad andmed, mida elektronkataloogi andmebaasis hoitakse, on bibliokirjed. Bibliokirjete põhisisu pole arvutiajastul muutunud võrreldes kaartkataloogide ja trükibibliograafiatega. Kirjed koostatakse ikka kirje- ja

kataloogimisreeglite järgi, kuid andmete arvutitöötlus ja infovahetus raamatukogude vahel seab suuremad nõuded kirjade korrektsusele ja reeglite ühtlustamisele.

Tänapäeval on peaaegu kõigis maades kataloogimisreeglite aluseks ISBDd (*International Standard Bibliographic Description*). ISBDd on IFLA poolt välja töötatud standardite rühm, mis määrab eri teavikulaadide kirjeldamisel esitatavate andmete valiku, vormi ja järjekorra, samuti kirjemärgid. ISBDd ei käsitle pealdisi, sisukirjelduselemente ega muud täiendavat teavet, mis kataloogikirjes antakse. Teisalt on eri maade kataloogimisreegleid *de facto* standardina mõjutanud angloameerika reeglid (AACR, *Anglo-American Cataloguing Rules*). Nende teine, uuendatud väljaanne AACR2 põhineb kirje osas samuti ISBD-l, kuid sisaldab lisaks pealdiste moodustamise reegleid. Seega on praeguseks maailmas saavutatud märkimisväärne ühtsus teavikute kirjeldamisel ja otsielementide määramisel.

Kirjade juurde kuuluv sisukirjeldus pole kaugeltki nii hästi reeglustatud, selles valitseb maailmas suur kirevus ja standardiseerimine on raske ülesanne. Kõige üldisemalt jaguneb intellektuaalne (inimese tehtav) indekseerimine liigituseks ja märksõnastuseks, mille tulemusel lisatakse bibliokirje formaalsele osale liigiindeksid ja märksõnad.

Ühtsed kirjereeglid on loodud silmas pidades bibliokirjade levitamist paberkandjal. Kirjade arvutitöötlus esitab nende vormile ja struktuurile oma nõuded, mis johtuvad osalt arvuti tööpõhimõtetest, osalt neist funktsioonidest, mis masinloetaval bibliokirjel täita tuleb.

Kaartkataloogimise ja elektronkataloogimise alused on mitmes suhtes erinevad. Kaartkataloogimise eesmärgiks on "valmistoote" – kataloogikaardi koostamine, elektronkataloogimisel ainult määratletakse bibliograafilised andmed ja varustatakse need tunnuskoodidega, sest elektrooniline kirje peab olema arvutisüsteemile täiesti ühemõtteliselt esitatud. Iga teaviku kohta on vaja koostada vaid üks põhikirje, millest arvuti abil võib saada mitmesuguseid "tooteid" – erineva pealdisega kataloogikaarte, bibliograafiaid, nimestikke, inventariraamatuid.

Elektrooniliste kirjade koostamisel poleks erilist mõtet, kui neid ei saaks andmekandjal või sideliinide kaudu levitada. Seepärast kehtib juba 1981. aastast standard ISO 2709 „Format for bibliographic information interchange on magnetic tape”, mis on kõigi arvutipõhiste kataloogimisüsteemide nurgakivi. ISO 2709 määrab kindlaks kirje põhistruktuuri ja mitmesuguste tehniliste andmete esitusviisi. Kirje täpsem struktuur ja kasutatavad koodid määratletakse nimetatud standardi rakendustes – bibliograafilistes vormingutes, millest tuntuimad kuuluvad MARC-rühma.

Normandmefail luuakse selleks, et hõlbustada normandmete ohjet ja kasutamist kataloogimisel. Normimist vajavad kirje tähtsaimad otsielemendid (pealdised): isikunimed, kollektiivinimed, ühtluspealkirjad, märksõnad, tihti ka liigiindeksid. Pealdise normkuju abil koondatakse kokku kõik antud autori või kollektiiv autori teosed, samuti nt klassikalise teose eri väljaanded, vaatamata sellele, millist kuju on kasutatud konkreetsetes teavikus. Variantvormid hoitakse normikirjes ära viidena. Märksõnade ja liigiindeksite normimisega saab parandada sisu indekseerimise järjepidevust. Seda tööd on tehtud ka kaartkataloogimise ajal, kuid arvuti pakub siingi palju rohkem võimalusi kui abikartoteegid.

Normandmed on soovitatav vormindada nii, et neid saaks kasutada bibliokirjade koostamiseks ja infootsinguks. Selleks on loodud omaette vormingud, mis ühilduvad bibliokirje MARC-vorminguga. Kui ühes ja samas andmebaasis on ühilduvates vormingutes bibliokirjed ja normikirjed, ei pruugi bibliokirje pealdist füüsiliselt sisaldadagi, vaid piisab normikirje kontrollnumbrist, mille abil pealdis bibliokirje vaatamise ajal nähtavaks saab.

Kolmanda andmekogumi, mida elektronkataloogi andmebaasis hoitakse, moodustavad registrid. Registrid tekitatakse bibliokirjade indekseerimise teel. See tähendab, et andmebaasihaldussüsteem koostab etteantud parameetrite järgi olulisematest kirjeandmetest loendid, mille abil ta hiljem leiab üles vajalikud kirjed. Registrite eesmärk on kiirendada otsingut, ilma nendeta peaks süsteem mingi tunnuse otsimisel läbi vaatama kõik kirjed, aga see võtaks vähegi suuremas andmebaasis liiga palju aega.

Kõige tavalisem registri vorm on nn pöördfail (*inverted file*), mis koosneb otsitunnustest ja viitadest, mis juhatavad antud tunnuseid sisaldavate kirjete juurde. Viitadeks võivad olla kirjete identnumbrid või aadressid, mis näitavad kirje asukohta andmebaasifailis. Seega sarnanevad elektronkataloogi registrid trükistes, nt raamatutes või bibliograafiates leiduvate registritega: raamaturegistris on aadressiks lehekülje number, bibliograafianimestikes kasutatakse viitadena tavaliselt kirjenumbreid. Vahe on selles, et elektronkataloogi registrid moodustatakse ja uuendatakse automaatselt.

Levinuimad registrid elektronkataloogides on autoriregister, pealkirjaregister, märksõnaregister, liigiregister ja võtmesõnaregister. Loendeid võib moodusteda ka teistest kirjeelementidest. Kõik sõltub sellest, milliseid otsitunnuseid peetakse antud süsteemi võimaluste piires otstarbekaks kasutajale esitada.

Samanimeliste registrite sisu võib süsteemiti erineda olenevalt sellest, millised MARC-vormingu väljad on indekseerimiseks valitud. Näiteks võib kollektiivautorid paigutada koos isikunimedega autoriregistrisse või moodustada neist omaette loendi. Isikunimed käsitlese ainena (personaalia) on mõnes kataloogis märksõnaregistris, teises jälle on moodustatud ühtne isikunimeregister, mille kaudu saab korraga kätte nii autori poolt loodud kui ka tema kohta kirjutatud teosed. Pealkirjaregister võib sisaldada nii üksikteoste kui ka jadaväljaannete pealkirju, teisel võivad viimased olla eraldi loendiks.

Otsivõimalused, mida lugejale pakuvad autori-, pealkirja-, märksõna- ja liigiregistrid, on olemas või vähemalt teostatavad ka kaartkataloogis. Elektronkataloogis täiendab neid (või on kasutussageduselt tihti lausa esikohal) võtmesõnaregister, mis moodustatakse valitud kirjeväljadel asetsevate keeleliste ühendite üksikutest sõnadest. Kõige kasulikumad otsinguks on igasuguste pealkirjade sõnad, aga ka kollektiivinimedes, märkustes jms sisalduvad sõnad.

Registrid on vajalikud nii andmebaasihaldussüsteemile kirjete ülesleidmiseks kui ka kataloogi kasutajale oma päringu väljendamiseks otsiterminite abil.

Sellise sisuga andmebaasi pidid hakkama koostama ELNET Konsortsiumi liikmesraamatukogud. Kui ISBD reeglid olid Eestis küllalt hästi tuntud ja ka kirjepealdiste osas polnud eri raamatukogude praktikas väga suuri erinevusi, siis MARC-vorming oli kataloogijatele seni üsna tundmatu. Formaalkataloogimisega seotud isiku- ja kollektiivnimede normandmestikku oli vaja alles looma hakata. Asusin INNOPACi evitusperioodil tööle Rahvusraamatukogusse ja normandmed sai üheks minu vastutusalaks, kuid käesolevas töös seda teemat otseselt ei käsitleta.

Minu ülesandeks konsortsiumi liigitamise ja märksõnastamise töörühma juhina sai INNOPACi sisukirjelduse ja aineotsingu meetodite planeerimine ja huvigruppide vahel kokkulepete saavutamine, et ära hoida kaost valdkonnas, kus see on väga kerge tekkima – suurte ühisandmebaaside hajutatud liigitamisel ja märksõnastamisel.

2. INDEKSEERIMISKEELED JA AINEOTSING

2.1 *Liigitamine ja märksõnastamine aineotsingu alusena*

Uurimused on näidanud, et üleminekul kaartkataloogilt elektronkataloogile aineotsingute osakaal võrreldes bibliograafiaotsingutega suureneb, moodustades kuni poole kõigist otsingutest (Markey 1984). Ja kõige keerulisem on kasutaja jaoks just aineotsing, mille aluseks on teaviku kirjele lisatud (või sellest automaatselt tuletatud) sisukirjelduselemendid – liigiindeksid, märksõnad, võtmesõnad.

Kaartkataloogis on liigitamine ja märksõnastamine alternatiivsed sisukirjeldusmeetodid. Sisukataloogi korraldamiseks on tulnud valida kas üks või teine. Elektronkataloogis on võimalik neid kasutada nii, et nad täiendavad teineteist, võttes arvesse mõlema eripära ning tugevaid ja nõrku külgi. Ideaalis peaksid liigi- ja märksõnaotsingu võimalused ning üleminek ühe rakenduselt teisele olema võrdselt hästi läbi töötatud.

Liigituse ja märksõnastuse põhilised erinevused on esitatud 9. artiklis.

1. Liigituse süstemaatiline lähenemine koondab materjali, võimaldab saada kogudest ülevaate erialade kaupa. Märksõnastuse alfabeetiline lähenemine hajutab materjali.
2. Liigitus on jäik. Uute teadusharude ja mõistete kajastamine liigitussüsteemides tekitab probleeme, sest nende uuendamine on aeglane protsess. Märksõnastus on paindlik. Uusi mõisteid saab süsteemi hõlpsasti lisada.
3. Liigitus võimaldab infootsijal paremini saada ülevaadet tema jaoks uuest ainealast. Märksõnastus sobib paremini konkreetsete päringute jaoks.
4. Liigituse ekspressiivne notatsioon võimaldab päringut hõlpsalt laiendada või kitsendada. Märksõnastikus mõistete hierarhilised suhted ei kajastu või kajastuvad vaid tesauruse pisihierarhiates.
5. Liigitus ei sõltu keelest. Märksõnastusega kaasnevad mitmesugused keeleprobleemid: kogude paljukeelsus, terminoloogia olukord, keele automaattöötuse keerukus, infootsija keeleoskus.
6. Liigitus opereerib pikkade abstraktsete liigiindeksitega. Märksõnastus kasutab loomuliku keele sõnavara.

- 7 Liigitussüsteem sobib enamasti ka teavikute paigutussüsteemiks. Märksõnad ei kõlba vähegi suuremate kogude paigutuse aluseks.
8. UDK, millega on harjunud Eesti raamatukogud, saame suhteliselt odavalt valmina kätte. Märksõnastikud tuleb meil endal välja arendada.
9. UDK liigitussüsteemi saab kasutada nii riigisiseses kui ka rahvusvahelises andmevahetuses. Märksõnastust on isegi Eesti piires raske standardida.

2.2. Märksõnastamine elektronkataloogis

2.2.1 Märksõnastamise meetodid

Mäletatavasti ei kujunenud Eesti teadusraamatukogudes elektronkatalooge juurutama asudes laiemat diskussiooni selle üle, kas märksõnastamine on vajalik ja missugust tüüpi märksõnastikud peaksid olema selle aluseks. Märksõnastamine elektrooniliste andmebaaside sisukirjeldusmeetodi ja aineotsingu võimaldajana tundus iseenesestmõistetav, kuigi seni oli kõigi universaalraamatukogude üldkatalooge üles ehitatud süstemaatiliselt UDK järgi. Siiski oli ka märksõnastamine üsna hästi tuntud – tõenäoliselt leidis igas raamatukogus mõni väiksem märksõnakataloog või -kartoteek. Märksõnastamise ajalugu Eestis on põhjalikumalt uurinud Marje Aasmets (Aasmets 2000, 2001). Mina olen koostanud lühikese ülevaate (6. artikkel) Tartu Ülikooli raamatukogu tegevusest märksõnastamise vallas kuni 1997 aastani, kus esitatakse andmeid ka eel-elektroonilisest ajastust.

Märksõnastiku tüübina leidis poolehoidu tesaurus. Kahes juhtivas raamatukogus – Eesti Rahvusraamatukogus ja Tartu Ülikooli raamatukogus – hakati peaaegu ühel ajal koostama universaalset tesaurust. Töötasin sel ajal ülikooli raamatukogus ja olin otseselt vastutav elektronkataloogi INGRID tesauruse kujundamise ja märksõnastamise juhendamise eest. Kuna selle valdkonna spetsialiste Eestis seni polnud (Rahvusraamatukogu kolleegid olid samasugused algajad iseõppijad nagu autorgi), tuli tesauruse ja märksõnastamise temaatikaga tutvuda eeskätt erialakirjanduse põhjal. Samas ei jäänud loetu pelgalt teoreetiliseks teadmiseks, vaid suur osa sellest sai praktika kaudu kriitiliselt läbi mõtestatud. Märksõnastamist, tesauruse olemust ja rakendusvõimalusi on käsitletud 4. ja 5. artiklis.

Jõudsin veendumusele, et traditsiooniline märksõnastus pole elektronkataloogis kuigi tõhus, ehkki enamik raamatukogusid on elektronkataloogile üle minnes jätkanud samade märksõnastike ja meetodite kasutamist mis kaartkataloogis. Sellel on kaks olulist põhjust: esiteks on uute, koordinatiivindekseerimiseks ja järelkoordineeritud otsinguks sobivate märksõnastike koostamine kulukas ja aeganõudev, ning teiseks – vahetades märksõnastikku ja metoodikat, tuleks kaartkataloogide retrokonversiooni käigus kõik teavikud ümber märksõnastada. Lugeses erialakirjandusest hinnanguid selle kohta, kui raske on lugejatele aineotsing Kongressi Raamatukogu märksõnastiku abil indekseeritud kataloogidest, tuli tõdeda, et elektronkataloog vajab teistsugust märksõnastamise metoodikat. Inversiooni, liitsõnade poolitamise, allmärksõnade, viidete jms abil hoolikalt kujundatud märksõnarubriigid ja -kompleksid ei ole infootsingul piisavalt etteaimatavad.

Võib öelda ka nii: kui kaartkataloogis on keskseks ühikuks rubriik, siis elektronkataloogis on selleks konkreetse teaviku kohta koostatud kirje. Seda ühte kirjet saab otsida mitmesuguste tunnuste järgi: autor, pealkiri, ilmumisaasta, ilmumiskoht, märksõna, liigiindeks. Põhimõtteliselt on kirje leitav iga selles esineva sümbolijärjendi (sõne) kaudu, seega pole aineotsinguks vaja tingimata märksõna või indeksit kasutada, vaid sobiva teose võib leida ka näiteks pealkirjas esineva sõna või sõnaosa järgi. Nn võtmesõna- ehk vabasõnaotsing on eriti levinud artiklite referaatandmebaasides, mis sisaldavad kirjutiste kokkuvõtteid, ja tekstiandmebaasides, kus elektroonilisel kujul on olemas kogu teaviku tekst. Enamik raamatukogukatalooge sisaldab siiski vaid traditsioonilisi kirjeelemente, mis võtmesõnaotsinguks eriti palju teksti ei paku, nii et sisu avamiseks tuleb kirjetele lisada täiendavat informatsiooni.

Kõige objektiivsem ja odavam oleks märksõnastada teaviku tekstis endas esinevate terminite abil. See on vaba märksõnastamise üks variant. Aga see teeks info taasleidmise raskeks: ammendava tulemi saamiseks peaks läbi proovima mitu sünonüümset väljendit mitmes keeles. Mittetekstiliste teavikute, nagu piltide ja helisalvestiste sisu tuleb igal juhul verbaalsesse keelde tõlkida. Lühidalt – infotehnoloogia praegusel arenguetapil ei saa me loobuda ohjatud märksõnastamisest raamatukogus hoitavate teavikute mitmekesisuse ja kataloogikirje iseloomu tõttu.

Kõige paindlikumaid otsivõimalusi pakub meetod, kui teaviku sisu väljendada rühma iseseisvate lihtmõisteid tähistavate loomuliku keele terminite kaudu (koordinatiivindekseerimine). Siis pole märksõnastamine enam teatud hulgale kirjetele sobiva ühise pealdise leidmine (vrd ingl k *subject heading!*), vaid iga teavik võib saada kordumatu, tema sisu kõige täpsemalt avava märksõnaloendi. Antud indekseerimiskeeles lubatud sõnavara on võimalik esitada erineval viisil – lihtsast märksõnaloendist põhjalikult läbitöötatud tesauruseni.

2.2.2. Tesauruse olemus ja ülesehitus

Terminiga *thesaurus* on erialakirjanduses üsna pillavalt ümber käidud: selleks on nimetatud nii kõige lihtsamat deskriptoriloetelu kui ka näiteks intelligentse süsteemi leksikaal-semantilist teadmusbaasi. Seetõttu on kohane esitada tesauruse määratlus selles tähenduses, milles seda terminit kasutatakse käesolevas töös. Selle järgi on *thesaurus* märksõnastik infotöötamise ja -otsingu jaoks, mis sisaldab loomulikust keelest valitud sõnu ja sõnaühendeid ning nendevahelisi semantilisi seoseid.

Tesauruse idee võtsid infoteadlased üle leksikograafiast. Leksikograafias nimetatakse tesauruseks sellist tüüpi sõnaraamatut, kus sõnad on korrapärased sisuliselt, mitte alfabeetiliselt. Ühte on koondatud kindla mõiste või ideega seotud sõnad ja väljendid. Keeletesauruste ülesanne on rikastada ja täpsustada inimese keelekasutust, aidata tal luua hea stiiliga tekste. Sõnaraamat juhib meid “ideesõnadelt” – mõistetelt, millest tahame rääkida või kirjutada – tekstisõnadeni, mis väljendavad meie mõtet kõige paremini. Teavikute töötlemisel infosüsteemides on vastupidine eesmärk: sisult sarnastele, kuid erinevaid sõnu sisaldavatele tekstidele tuleb leida mingi ühine “ideesõna”. Kuigi need kaks protsessi on vastupidise suunaga ning keeletesaurused ei saa asendada infotesaurusi ja vastupidi, on nad piisavalt sarnased selleks, et abivahendid nende protsesside jaoks võiksid kanda ühist perekonnanime.

Idee rakendada tesaurust infootsingus ja esimesed katsetused selles vallas pärinevad 1950. aastate teisest poolest. Pikemalt olen tesauruse arenguloost kirjutanud 4. artiklis.

Tesaurused on ainukesed indekseerimiskeeled, mille sõnavaravalik, struktuur, esitus ja haldus on rahvusvaheliselt standarditud. Nende kohta käivad standardid ISO 2788:1986 “Documentation. Guidelines for the establishment and development of monolingual thesauri” ja ISO 5964:1985 “Documentation. Guidelines for the establishment and development of multilingual thesauri” Olen neid standardeid analüüsinud 8. artiklis. Enamik riiklikke standardeid lähtub ISO standarditest, kuid arvestab ka oma maa traditsioone ja keele eripära. Üldtunnustatud käsiraamat, mis toetub kehtivatele standarditele, kuid annab tesaurusega seotud probleemidest palju avarama pildi, on J. Aitchisoni ja A. Gilchristi “Thesaurus construction and use: a practical manual” (4. tr. 2000).

Tesauruse struktuuri moodustavad terminid ja nendevahelised seosed. Terminid väljendavad märksõnastiku ainevaldkonna olulisi mõisteid. Loomulikule keelele, ka oskuskeelele on omane sünonüümia: üht mõistet saab väljendada mitut moodi. Tesauruse ülesanne on piirata infosüsteemis kasutatavate terminite hulka ja üks meetod selleks ongi sünonüümsete väljendite kokkuviimine. Seega sisaldab tesaurus kahesuguseid termineid: märksõnad ja äraviiteterminid, mis järjestatakse tähestikuliselt ühtsesse nimestikku.

Tesauruses kajastub ka teatud hulk mõistetevahelisi semantilisi suhteid. Neid väljendatakse viidetega termini juurest teistele terminitele. Tavaliselt eristatakse kolme semantilist suhet: ekvivalentsisuhe, hierarhiasuhe ja assotsiatsioonisuhe.

Ekvivalentsiseos e samaväärsusseos moodustatakse selliste terminite vahele, mis tähistavad üht ja sama mõistet või niivõrd lähedasi mõisteid, et märksõnastamisel ja infootsingul ei ole otstarbekas neid lahus hoida. Üks ekvivalentide klassi liige valitakse märksõnaks (*preferred term, descriptor*), teine läheb tesaurusesse äraviiteterminina (*non-preferred term, entry term, non-descriptor*). Ühe märksõna juures võib olla mitu äraviiteterminit.

Hierarhiasuhe on selliste mõistete vahel, millest üks sisaldub tervikuna teises, moodustades sellest osa. Rangelt võttes ja standardi järgi on hierarhiaseos lubatud ainult selliste terminite vahel, mis kuuluvad samasse semantilisse kategooriasse, st tähistavad mõlemad kas nähtust, omadust, tegevust, teadusdistsipliini vms.

Hierarhiasuhte alaliigid on geneeriline e soo-liigisuhe, partitiivne e osa-tervikusuhe ja üldmõiste-üksikjuhtumi suhe. Geneerilise seose puhul on laiemaks soomõiste ja kitsamaks liigimõiste. Tavaliselt on ühel soomõistel mitu alluvat liigimõistet. Partitiivse suhte puhul väljendab laiem mõiste tervikut ja kitsam selle osa. Selliseid sõnarühmi, kus osa ja terviku suhe oleks püsiv hoolimata kontekstist, pole kuigi palju. Üldmõiste ja üksikjuhtumi suhtega (*instance relationship*) on tegemist siis, kui hierarhia moodustavad üldnimetus ja sellest kitsamate terminitena nimed.

Assotsiatsiooniseost on raskem defineerida kui eelmisi seoseliike. Üldiselt esineb assotsiatsiooniseos selliste terminite vahel, mis on mõisteliselt küll tihedalt seotud, aga mitte hierarhiasuhtes, samuti pole nad ühe ja sama ekvivalentide klassi liikmed. Assotsiatsiooniseoses termineid esineb rohkesti humanitaar- ja sotsiaalaladel, kus kategooriate ja hierarhiate moodustamisest on tihti parem loobuda, et vältida subjektiivseid otsustusi.

Tüüpiline märksõnaartikkel näeb tesauruses välja nii:

RAHANDUSPOLIITIKA

asendab terminit	<i>finantspoliitika</i>
laiem termin	MAJANDUSPOLIITIKA
kitsam termin	MAKSUPOLIITIKA
seotud termin	RIIKLIK RAHANDUS

Äraviitetermini puhul ei anta tavaliselt muud informatsiooni kui viide märksõnale:

finantspoliitika

kasuta	RAHANDUSPOLIITIKA
--------	-------------------

Eelkirjeldatud viisil esitatakse märksõnaartiklid sellises tesauruses, mille põhiosa on tähestikuline. Enamik tesaurusi ongi nii üles ehitatud. Süstemaatiline osa piirdub neis laiade temaatiliste rühmade kaupa loetletud deskriptoritega. Näiteks võib tuua “Eesti üldise märksõnastiku” Juhul kui süstemaatika on hästi läbi töötatud ja moodustab tesauruse tähtsama osa, jäetakse tähestikregistrist tavaliselt hierarhiaseosed välja. Selline on “Medical Subject Headings”

Olen osalenud kahe universaalse tesauruse koostamises ja haldamises. Aastatel 1994 – 1997 juhendasin Tartu Ülikooli raamatukogu elektronkataloogi INGRID tesauruse koostamist ja alates 1998. aastast vastutan “Eesti üldise märksõnastiku” eest. Ligi kümneaastane kogemus lubab väita, et ISO standardite ja tesauruse koostamise käsiraamatu ranged nõuded terminite valikule ja semantiliste seoste moodustamisele on universaalsete tesauruste puhul raskesti täidetavad. See ilmnes juba INGRIDi tesauruse puhul ning mõned probleemid olen välja toonud 5. artiklis.

Tesauruse standardid lähtuvad eeldusest, et tegemist on erialamärksõnastikuga sõltumatu erialaandmebaasi jaoks, mis sisaldab teaduspublikatsioonide kirjeid. Seega pärineb korrastatav terminoloogia ühest oskuskeelee allkeelest. Sellisel juhul on võimalik loogilise analüüsi abil see terminoloogia standardseoste abil korralikult süstematiseerida. Universaalraamatukogu kataloogi sisu on palju mitmekesisem, hõlmates eri ainevaldkondade teadus-, aime- ja tarbekirjandust, ilukirjandust, erilaade (helisalvestised, noodid, graafika- ja kaarditeavikud), käsikirju ja arhivaale paljudes keeltes ja eri ajastutest. Ühtse mõistesüsteemi ehitamine sellise teadmiste universumi indekseerimiseks on äärmiselt raske, seetõttu ei saa ka üldine tesaurus olla väga range struktuuri ega stiilipuhta sõnavalikuga. Raamatukogukataloogi märksõnastamisel on paratamatult vaja kasutada rohkem liitmõisteid (s.t eelkoordineeritud mõisteid) kui erialaandmebaasides, et muuta märksõnade tähendusi täpsemaks ja hoida erialad lahus. Ranged semantilised piirangud hierarhiaseoste moodustamisel võivad halvendada tesauruse tõhusust infootsingus, kui kasutaja oma infovajadusele vastavaid märksõnu valib. Neid asjaolusid on Eestis koostatud tesauruste ülesehituses püütud arvestada, kuid põhiolemuselt vastavad need ISO standarditele.

2.2.3. Tesaurus ja arvuti

Tesauruse koostamine ja edasine haldamine sisaldab rohkesti mehhaanilist, kuid samas suurt täpsust ja inimeselt palju aega nõudvat tööd. Seepärast on tänapäeval üldine, et tesaurusi tehakse ja toimetatakse arvuti abil. Kasutatav tarkvara põhineb tavaliselt mõnel universaalsel andmebaasihaldussüsteemil, mille tesaurust haldav asutus on kohandanud vastavalt oma vajadustele ja võimalustele. Leidub ka

kommertspakette, mis on mõeldud spetsiaalselt tesauruse jaoks, kuid eesti keelt toetavaid nende hulgas pole. Eesti tesauruste tarkvara on projekteeritud kohapeal.

Kõige raskem on tesauruse juures seoste moodustamine ja kontroll. Muidugi ei suuda arvuti hinnata, kas seos on semantiliselt korrektne. See jääb inimese ülesandeks. Kuid sobiva tarkvara abil saab vältida vigu ja võita aega seoste vormistamisel: piisab, kui sisestada vaid seose üks pool, teine pool tekib automaatselt ja midagi ei unustata ära. Kahepoolsete seoste genereerimine ühepoolse sisendi põhjal on miinimumnõue, millele peab vastama tesauruse koostamise tarkvara (Milstead 1993, 390). Aga sellest ei piisa paberitööst loobumiseks. Kui juba arvuti abiks võtta, peaks programm võimaldama säilitada kogu tesauruse teksti, k.a märkused ja definitsioonid, sortida ja järjestada sõnu mitmesuguste tunnuste alusel, teha väljatrükke kogu tesaurusest või selle osast. Kui eesmärgiks on tesaurus publitseerida, peab olema võimalik kujundada andmebaasist hea tüpograafilise kvaliteediga fotoladu.

Masinloetav tesaurus on suurepärane abivahend märksõnastamisel, eriti kui tesaurus ja bibliograafiaandmebaas on omavahel integreeritud. Hea programm ei lase märksõnastada terminitega, mis pole tesauruses autoriseeritud (nii antakse märku ka kirjavigadest). Samm edasi on võimalus valida kirjele märksõna ekraanile tellitud nimestikust ühe klahvilöögiga. Kui märksõnastaja juhtub valima äraviitetermini, asendub see automaatselt vastava deskriptoriga. Kasulik on näha ka terminitevahelisi seoseid, et leida kõige õigemad märksõnad. Sellise märksõnastussüsteemi eelis on vigade vältimine, kuid programmeerimisel tuleb hoolitseda, et terminite leidmine tesaurusest ja automaatne kirjele lisamine ei võtaks kokkuvõttes rohkem aega kui nende väljakirjutamine.

Andmebaasiga integreeritud tesaurus võimaldab parandada märksõna nii, et muutus peegeldub kohe kõigi kirjete juures, mis on selle märksõnaga indekseeritud. See on eriti tähtis juhul, kui tesauruse loomine ja märksõnastamine toimuvad paralleelselt. Siis tuleb sageli parandusi teha. Ka hiljem on vaja asendada vananenud termineid uutega. Kui mingi sõna tesaurusest kustutada, peab see kustuma ka kõigi kirjete juurest. Aeganõudvam, kuid turvalisem on selline variant, kus kustutada saab ainult kasutamata märksõnu, s.t vastavad kirjed tuleb enne üle vaadata ja parandada,

eemaldades soovimatu märksõna. Alles siis saab märksõna tesaaurusest kustutada või muuta äraviiteterminiks.

INGRIDi tesaaurusemoodul vastas hästi ülalkirjeldatud mudelile. Mooduli programmeeris Ingrid Saare, kes on ka “Eesti üldise märksõnastiku” praeguseni kasutatava haldusprogrammi autor. Täpsemalt on INGRIDi tesaauruse funktsioneerimisviisi kirjeldatud 5. artiklis.

Tesaauruse ühendamine integreeritud raamatukogusüsteemiga oli 1980. aastate teisel ja 1990. aastate esimesel poolel maailmas üks ideesid, mille abil loodeti täiustada aineotsingut elektronkataloogis (vt nt Congreve 1986; Markey 1988; Norgard jt 1993; O’Brien 1994). Sellel taustal tuleb tunnustada ülikooli raamatukogu edumeelsust. Tolleaegsetest kommertsraamatukogusüsteemidest, mis võimaldasid tesaauruse koostamist ja kasutamist, oli vaid mõnes vastav moodul täiuslikum kui INGRIDis.

INNOPACi süsteem ei sisalda tesaauruse moodulit, kuna lähtub angloameerika kataloogimistraditsioonist, milles sisukirjelduse aluseks on eelkoordinatsioonil põhinev Kongressi Raamatukogu märksõnastik. Osaliselt täidavad integreeritud tesaauruse ülesannet märksõnade normikirjed, kuid need on killustatud märksõnaregistrisse ning pole lugejakataloogis sealtkaudugi tervikuna kättesaadavad. Kõige olulisem tesaauruse funktsioon – semantiliste seoste eksponeerimine – INNOPACis siiski osaliselt (äraviidete ja sideviidetena) toimib.

2.3. Universaalse kümnendliigituse (UDK) rahvusvahelisest levikust

2.3.1. UDK rakendusala

Eesti raamatukogudes oli 1990. aastate alguses tunda kõhklust UDK-liigituse jätkamise suhtes. Suurte raamatukogude liigikataloogidest oli raske vajalikku kirjandust üles leida, tulevaste elektronkataloogide kohta arvati, et neist nagunii keegi indeksi järgi otsima ei hakka, liigitustabelid olid „nõukogude” ja vene keeles, Põhjamaade kolleegid pidasid UDKd aegunud ja väljasurevaks nähtuseks, selle rahvusvaheline haldamine oli käest ära lastud jne. Siiski otsustas Eesti

Raamatukoguhoidjate Ühingu liigitamise ja märksõnastamise toimikond 1993. aastal hakata ette valmistama esimest omakeelset UDK väljaannet.

Peale liigitustabelite tõlkimise oli oluline süveneda liigitussüsteemide toimimisse elektroonilises keskkonnas. Sel teemal kirjutasin oma diplomitöö „UDK kasutamine arvutikataloogides ja andmebaasides” (1994). Hiljem olen UDK temaatikat edasi uurinud ja avaldanud tulemused 7 artiklina. Artiklis käsitletakse UDK rakendusvaldkondi ja levikut eri maades.

Sünteesliigitusena on UDK struktuurilt väga paindlik. See annab võimaluse kasutada teda mitmeks otstarbeks, valides vajaliku detailsuse astme ning sünteesiviisi.

Enamikus maailma raamatukogudes on liigitussüsteemid kogude paigutuse aluseks. UDK on levinud rohkem reaalteaduslikes ja tehnikaraamatukogudes. Suur universaalkogu on UDK järgi paigutatud näiteks Pompidou Keskuse raamatukogus Pariisis. Paigutuseks tavaliselt UDK süsteemi lihtsustatakse.

Kui suurem osa kogusid paikneb kinnistes hoidlates, on raamatukogu teatmestu tähtsaks osaks süstemaatiline kataloog. UDK on ainuke üldtunnustatud liigitussüsteem, mis on loodud just süstemaatilise kataloogi vajadusi silmas pidades (teised on kujundatud eelkõige paigutussüsteemidena). Bibliokirjetes kasutatakse detailsemaid indekseid, kuid ka siin saab liigitustaset varieerida vastavalt kogude suurusele ja liigitatavate teavikute iseloomule. Süstemaatiline kataloog on eelkõige Euroopa nähtus.

Elektronkataloogile üle minnes on need raamatukogud, kes varem pidasid süstemaatilist kataloogi UDK järgi, tavaliselt jätkanud indekseid lisamist kirjetele. Ehkki lugejad eelistavad konkreetsete päringute jaoks võimaluse korral märksõna- või võtmesõnaotsingut, on liigitus vajalik kataloogi sisust süstemaatilise pildi saamiseks. Sobiva tarkvara puhul on UDK elektronkataloogi sirvimiseks parim vahend. Liigituse potentsiaalile aineotsingus juhtis tähelepanu juba Svenonius (1983). Enamik kommertsiaalseid raamatukogusüsteeme, kaasa arvatud INNOPAC, suudavad liigiindekseid töödelda kui tavalisi sümbolijadasid, s.t neid salvestada, järjestada, kuvada, indekseerida ja otsida. UDK-otsingu põhiprobleem on selles, et tavakasutaja

ei valda UDK notatsiooni. Euroopas on välja arendatud mõned raamatukogusüsteemid, milles UDK indeksid ja märksõnad on omavahel automaatselt seotud (Buxton 1993). Kuid minu arvates leiab liigitussüsteemide kasutamine aineotsinguks elektronkataloogides vaid siis laiemat kõlapinda, kui tehnilised lahendused saavad tugineda piisavalt detailsetele standardsetele andmevormingutele. McIlwaine (2000) on analüüsinud MARC-vormingute liigitusandmete osa puudusi ja esitanud täiendusettepanekuid, samuti rõhutanud liigiindeksite normandmefailide osatähtsust liigiotsingu võimaluste kujundamisel.

UDKd on kasutatud infoväljaannete koostamisel kui materjali korrastamise vahendit. Siin tuleb kasuks UDK detailsus, sest artiklid, teadusaruanded, konverentsiettekanded jms on tavaliselt kitsamatel teemadel kui monograafiad. Infotalitustes on kergem minna kaasa süsteemis tehtavate muudatustega, sest mahukat ümberliigitamist nagu raamatukogus need kaasa ei too: regulaarselt ilmuva infoväljaande järgmises numbris kasutatakse lihtsalt uuendatud süsteemi.

Trükitud infoväljaandeid on asendamas sidusandmebaasid. Nendes kasutatakse indekseerimiseks ja infootsinguks tavaliselt deskriptoreid, kuid vahel ka UDK indekseid. 1985. aastal leidis 39 Austraalias, Austrias, Hispaanias, Hollandis, Kanadas, Saksamaal, Soomes, Suurbritannias, Šveitsis, Taanis ja USAs toodetud andmebaasi, mille otsikeeleks on UDK.

Laialt on levinud normdokumentide märgendamine UDK indeksiga, seda eelkõige ISOga (*International Organization for Standardization*) ühinenud maades, aga ka mujal. 1985. aastal oli neid riike 47 (tolleaegse poliitilise kaardi järgi), lisaks kasutasid UDKd oma standardites regionaalsed ja rahvusvahelised organisatsioonid.

Küllalt levinud on UDK kasutamine rahvusbibliograafiates. IFLA on seoses UBC (*Universal Bibliographic Control*) programmiga soovitanud struktureerida rahvusbibliograafia väljaanded süstemaatiliselt mõne üldlevinud liigitussüsteemi järgi. Enamasti on selleks saanud DDC (*Dewey Decimal Classification*). 1993. aasta andmetel kasutati UDKd rahvusbibliograafias 21 riigis. See arv pole kindlasti enam täpne, sest arvesse polnud võetud kõiki NSV Liidu ja Jugoslaavia lagunemisel tekkinud riike.

Rahvusbibliograafiates leiduvad UDK indeksid on tavaliselt väga üldised. See ei häirinud perioodiliselt ilmuvates trükiväljaannetes, kuid elektronandmebaasides, kus andmed pidevalt kumuleeruvad, ei võimalda liiga üldine liigitus enam tulemuslikku otsingut.

UDK uusim rakendusvaldkond on Interneti inforessursside temaatiline korrastamine. Mitmeid näiteid selle kohta leiab Suurbritannia infoserveritest.

2.3.2. UDK levik maailma maades

Kõige rohkem on andmeid leida Euroopa riikide kohta. Suurbritannial on juhtiv positsioon ingliskeelsete UDK tabelite väljaandmises ja levitamises. *British Standards Institution* on UDK Konsortsiumi asutajaliige. 1980. aasta andmetel kasutas 22% Suurbritannia raamatukogudest ja infoasutustest UDKd. Saksamaal tegeles 1990. aastani UDK haldamisega *Deutsches Institut für Normung*. 1982. a. andmetel oli Saksamaal 167 UDKd kasutavat asutust. Taanis ja Rootsis kasutatakse UDKd eelkõige teadus- ja erialaraamatukogudes, neis keeltes on olemas lühendatud tabelid. Soomes oli UDK-l tugev positsioon alates 1960. aastatest. Mitmesuguste bibliograafiaandmebaaside enamik kirjeid on seal liigitatud UDK järgi. Lühendatud soomekeelseid tabeleid on välja antud mitmes trükis. Portugalis on UDK valdav raamatukoguliigitus. Raamatukogude koondkataloogis PORBASE on see esmane ja kohustuslik sisukirjeldusvahend. Lääne-Euroopa riikidest on UDK küllalt hästi tuntud veel Belgias, Hispaanias, Hollandis, Itaalias, Norras, Prantsusmaal. Portugali- ja hispaaniakeelsed tabelid on mõjutanud UDKd rakendama ka Ladina-Ameerika riike.

Kesk- ja Ida-Euroopa maades on liigituspoliitika ja -praktika olnud enam tsentraliseeritud kui Lääne-Euroopas. UDK on liigitussüsteemina domineerinud Ungari, Tšehhi, Slovakkia, Poola jt maade raamatukogudes. Vastavates keeltes on avaldatud mitmesuguseid tabeleid. On tähelepanuväärne, et 1995. aasta sügiseni UDK Konsortsiumi poolt müüdud 20 publitseerimis litsentsist 11 olid omandanud endised sotsialismimaad.

Aasia ja Aafrika maade raamatukogud kasutavad laialdaselt liigitussüsteeme, mis on koostatud Läänemaades, eriti Dewey kümnendliigitust ja UDKd. Need on küll Lääne kultuuri kesksed ega arvesta vajalikul määral Ida tsivilisatsioonide omapära, kuid universaalsete liigitussüsteemide väljatöötamine on küllalt kallis ning arengumaadel ei jätku selleks raha ega asjatundjaid. Ainult sellised rikkad riigid nagu Jaapan või suured riigid nagu Hiina on võimelised ülal pidama oma rahvuslikke liigitussüsteeme.

On avaldatud andmeid, et UDK-l on 100 000 institutsionaalset ja individuaalset kasutajat umbes 60 riigis. Paljudes maades, eriti Ida-Euroopas, on see ühtseks liigitussüsteemiks kõigis või enamikus raamatukogudes. Tabelid on täielikus, keskmises või lühendatud mahus olemas vähemalt 26 keeles, kaasa arvatud eesti keel. Viimastel aastatel on ilmunud palju uusväljaandeid, mis annab tunnistust UDK jätkuvast vajalikkusest ja populaarsusest.

2.4. Indekseerimise ja indekseerimiskeelte standardimine

Sisukirjelduse standardimine on vastuoluline teema. Teavikute sisu mõistmises ja väljendamis, samuti aineotsingul oma infovajaduse formuleerimises mängivad subjektiivsed faktorid väga olulist rolli. Teisalt saab infootsingu tulemuste rahuldavat täielikkust ja täpsust saavutada ainult ühtse ja järjepideva sisukirjeldusega, mis tähendab teatud standardite ja reeglite järgimist. Standardimine on oluline ka infosüsteemide ühilduvuse ja koostalitlusvõime tagamiseks. Sisukirjelduse ühtlustamisest ja standardimisest rääkides on otstarbekas eristada andmebaasi tasandit, riiklikku (ühe maa) tasandit ja rahvusvahelist tasandit. Võimalused saavutada järjepidevat sisukirjeldust on andmebaasi tasandil mõõtnatult suuremad kui näiteks rahvusvahelisel tasandil, kuigi ka andmebaasid on väga erineva ea, suuruse ja koostajate arvuga.

Käesoleva magistritöö tarbeks olen uurinud eelkõige rahvusvahelisi standardeid, milles käsitletakse indekseerimist ja indekseerimiskeeli. Tulemused on esitatud 8. artiklis.

Indekseerimisprotsessi, eriti selle esimesi, dokumendi sisu kontseptuaalse analüüsiga seotud etappe, on Williamsoni (1995) arvates peaaegu võimatu standardida. Siiski pakub ISO 5963:1985 „Documentation. Methods for examining documents, determining their subjects, and selecting indexing terms” mõningaid kasulikke üldisi lähtekohti. Standardis nimetatakse dokumentide olulisemad osad, millega tuleb kindlasti tutvuda, hoiatatakse ainult pealkirja järgi indekseerimise eest, rõhutatakse, et mittetekstiliste dokumentide puhul tuleb rakendada erimetoodikat jms. Indekseeritavate mõistete valikul tuleb alati hinnata nende potentsiaalset väärtust dokumendi ainestiku väljendamiseks ja dokumendi taasleidmiseks.

Standardis ei määrata, milline peaks olema indekseerimise ammendavus ja spetsiifilisus, vaid rõhutatakse, et ühegi dokumendi sisu kirjeldamiseks pole olemas ühte õiget terminikogumit. Ammendavuse ja spetsiifilisuse nõuded saab seega kehtestada ainult andmebaasi tasandil.

Indekseerimiskeelte hulka kuuluvate tesaauruste koostamist puudutavaid standardeid käsitlesin jaotises 2.2.2. Peale ISO on ka IFLA püüdnud anda oma panuse märksõnastike „hea tava” kujundamisse. IFLA liigitamise ja märksõnastamise sektsiooni töögrupi ettevalmistatud dokumendis „Principles underlying subject heading languages (SHLs)” (1999) loetletakse 11 printsiipi, millest tuleks lähtuda märksõnasüsteemide koostamisel ja rakendamisel, ning illustreeritakse neid põhimõtteid eri maade indekseerimissüsteemidest leitud näidetega. Kahjuks pole sealjuures printsiipe peaaegu üldse selgitatud, seetõttu on võimalik neid mitut moodi tõlgendada. Kuigi palju praktilist kasu sellest dokumendist ei saa.

Verbaalsete indekseerimiskeelte rahvusvahelist standardimist ja koostalitlust raskendab asjaolu, et keelte struktuur ja sõnavaras väljenduvad mõistesüsteemid võivad olla väga erinevad. Keelebarjääride ületamiseks rahvusvahelises infovahetuses ja mitmekeelses infootsingus tuleks keskenduda senisest enam mitmekeelsete tesaauruste, liigitussüsteemide ja vahekeelte standardite väljatöötamisele.

3. KOKKUVÕTE JA JÄRELDUSED

Magistritöös keskendusin järgmistele probleemidele.

Esiteks uurisin elektronkataloogide arengut ja kasutatavaid andmestruktuure, et mõista, millest tuleneb nende otsiomaduste mahajäämus võrreldes teiste infootsisüsteemidega. Püüdsin aru saada, millised aineotsingu võimalused üldse on põhimõtteliselt teostatavad meile kättesaadavates raamatukogusüsteemides.

Teiseks püüdsin leida elektronkataloogile kõige sobivamaid indekseerimiskeeli, sisukirjelduse meetodeid ja nende alusel kujundatavaid aineotsingu võimalusi. Selleks võrdlesin eri aspektidest traditsioonilisi indekseerimisviise liigitamist ja märksõnastamist; analüüsisin erinevaid märksõnastamise meetodeid, eelkõige koordinatiivindekseerimist ja eelkoordineeritud märksõnastamist; uurisin tesauruse ülesehitust, koostamisvõtteid ja integreerimist bibliograafilise andmebaasiga; uurisin UDK seisundit ja tulevikuperspektiive; selgitasin välja indekseerimist puudutavad rahvusvahelised standardid ja analüüsisin nende rakendatavust ning kasutegurit praktilises töös.

Töö eesmärgiks oli leida Eesti raamatukogudele sobivad praktilised mudelid sisukirjelduse ja aineotsingu jaoks elektronkataloogides.

INGRIDI kataloogi jaoks ei suutnud ma terviklikku mudelit välja töötada, sest süsteemi juurutamise algetapil nappis teadmisi ja kogemusi, mõne aasta pärast aga selgus, et Eestis on perspektiivikam evitada ühisandmebaasidel põhinev võimas integreeritud raamatukogusüsteem.

Kuna INGRID oli ühe raamatukogu süsteem ning sisukirjeldus ja tesauruse koostamine olid koondatud ühte allüksusesse, oli võimalik programmeerija, indekseerijate ja toimetaja koostöös kujundada peamiseks sisukirjelduse ja aineotsingu vahendiks integreeritud tesaurus.

ELNET Konsortsiumi ühiskataloogide jaoks see idee ei sobinud. Esiteks ei võimalda INNOPAC tehniliselt sellist tesaurust süsteemiga liita, teiseks poleks intellektuaalselt võimalik kõigi konsortsiumi liikmesraamatukogude kogude sisu ühe tesaurusega hõlmata, kolmandaks eeldaks keskne range tesaurus märksõnastamise sellist ühtsust ja järjepidevust, mida eri suunitlusega raamatukogude vahel ei ole võimalik saavutada.

ESTERi jaoks kujunes mudel, mille põhiprintsiibid on järgmised.

1. Sisukirjelduseks tuleb kasutada kaht teineteist täiendavat meetodit – liigitamist ja deskriptoriga märksõnastamist. Liigitus võimaldab andmebaasi sisu süstemaatilise liigenduse ja pakub infootsinguks eelkoordineeritud mõisteid. Verbaalsed deskriptorid (märksõnad) annavad võimaluse mõistete vabaks järelkoordinatsiooniks otsingufaasis, nii omavahel kui ka kombinatsioonis võtmesõnadega.
2. Liigitamiseks ja märksõnastamiseks tuleb kasutada ühtseid indekseerimiskeeli – UDK eestikeelset väljaannet ja „Eesti üldist märksõnastikku” Teatud ulatuses on vältimatu ka vaba märksõnastamine.
3. Kasutajale tuleb sobivate otsitunnuste leidmiseks pakkuda maksimaalset abi sirvitavate registrite, integreeritud normandmefailide ja juhenditena.
4. Maksimaalselt tuleb kasutada süsteemi pakutavaid autoriteetsuskontrolli võimalusi ja pidevalt jälgida raamatukogude liigitus- ja märksõnastuspraktikat. Keskne roll peab selles olema kõiki liikmesraamatukogusid hõlmaval liigitamise ja märksõnastamise tööühmal.

Tuleb tunnistada, et mudeli aluseks on peale asjakohase uurimistöö olnud usk ja lootus, et Eesti-suuruses riigis on koostöö ja koordineeritusega võimalik tagada ühiskataloogide sisukirjelduse ühtsus ja järjepidevus palju suuremal määral kui teistes Euroopa riikides.

Käesoleva magistritöö ülesandeks ei olnud käsitleda kirjeldatud mudeli elluviimise käiku ega hinnata selle edukust. Kindlasti ei ole kõik mudeli komponendid täielikult realiseeritud või ei toimi need soovitud efektiivsusega. Esitan mõned edasise uurimis- ja arendustöö probleemid.

Milline on ESTERi kasutaja otsikäitumine? Kas ta oskab oma päringut indekseerimiskeeles väljendada? Kas ta teeb vahet eri registritel ja otsitunnustel? Kas ta valib otsinguks sobiva registri? Kas ta oskab kasutada viiteid ja muid abivahendeid?

Kuidas jõuda selleni, et Eestis ei arendataks paralleelselt kahte universaalset tesaurust, „Eesti üldist märksõnastikku” ja Tartu Ülikooli raamatukogu tesaurust? Kas neid oleks võimalik ühendada? Kes oskaks ja tahaks seda teha? Kui kaua see aega võtaks?

Kas ESTERis on võimalik välja arendada kasutajasõbralik süstemaatiline otsing UDK indeksite ja registri alusel? Kas meil on põhjust loota seda omanikfirmalt? Kas me saaksime ise midagi ära teha?

Tõenäoliselt veeretab raamatukogude infotehnoloogiline areng meie ette veel selliseid probleeme, mida praegu ei oska ette näha.

KASUTATUD KIRJANDUS

Aasmets, M. (2001). Märksõnastamine Eesti raamatukogunduses nõukogude perioodil. *Eesti Akadeemiline Raamatukogu. Aastaraamat 2000*. Tallinn, lk. 168-179.

Aasmets, M. (2000). Märksõnastamise areng Eesti raamatukogunduses 1918-1940. *Eesti Akadeemiline Raamatukogu. Aastaraamat 1999*. Tallinn, lk. 125-135.

Buxton, A. B. (1993). Computer searching of UDC numbers. *Encyclopedia of library and information science 51* New York: Marcel Dekker, pp. 132-151.

Congreve, J. (1986). Problems of subject access: (i) automatic generation of printed indexes and online thesaural control. *Program*, 20(2), 204-210.

Hildreth, C. R. (1987). Beyond Boolean: designing the next generation of online catalogs. *Library Trends*, 45(4) (Spring 1987), 647-667

Hildreth, C. R. (1995). Online catalog design models: are we moving in the right direction? A report submitted to the Council on Library Resources [uuendatud 2000]. <http://phoenix.liu.edu/~hildreth/clr-opac.html> (18.05.04)

Hjørland, B. (1997). Information seeking and subject representation: an activity-theoretical approach to information science. London: Greenwood Press

Kruus, A. (2000). Liigitamine ja märksõnastamine abivahendina andmebaasidest informatsiooni otsingul. *Eesti Akadeemiline Raamatukogu. Aastaraamat 1999*. Tallinn, lk. 118-124.

Larson, R. R. (1991). Between Scylla and Charybdis: subject searching in the online catalog. *Advances in Librarianship*, 15, 175-236.

Markey, K. (1984). Subject searching in library catalogs: before and after the introduction of online catalogs. Dublin: OCLC.

Markey, K. (1988). Integrating the machine-readable LCSH into online catalogs. *Information Technology and Libraries*, 7(3), 299-312.

McIlwaine, I. C. (2000). The Universal Decimal Classification: a guide to its use. Hague: UDC Consortium.

Milstead, J. L. (1993). Thesaurus management software. *Encyclopedia of library and information science 51*. New York: Marcel Dekker, pp. 389-407

Norgard, B. A., Berger, M. G., Buckland, M., Plaunt, C. (1993). The online catalog: from technical services to access service. *Advances in Librarianship*, 17, 111-148.

O'Brien, A. (1994). Online catalogs: enhancements and developments. *Annual Review of Information Science and Technology*, 29, 219-242.

Su, S.-F (1994). Dialogue with an OPAC: how visionary was Swanson in 1964? *Library Quarterly*, 64(2), 130-161.

Svenonius, E. (1983). Use of classification in online retrieval. *Library Resources and Technical Services*, 27, 76-80.

Svenonius, E. (2000). The intellectual foundation of information organization. Cambridge: MIT Press.

Williamson, N. J. (1995). Standards and standardization in subject analysis systems: current status and future directions. *Subject indexing: principles and practices in the 90's*. München: Saur, pp.278-291.

ONLINE CATALOGUES AND SUBJECT ACCESS: DEVELOPMENT PROBLEMS IN ESTONIA IN 1990s

Summary

Keywords: library automation, OPAC, subject access, indexing languages, UDC, thesaurus, subject indexing, standardisation

The implementation of online catalogues and other online bibliographical databases began in Estonian libraries in the beginning of 1990s. In the context of electronic cataloguing, the methods of subject description were one of the most important areas that had to be reasoned. For decades Estonian libraries had used classification as the main subject description method. Though verbal subject indexing was practised, there were no general subject heading lists and methods of subject indexing varied a lot. Library practices in developed countries and prevailing standpoints in information science about the methods of subject description and subject searching in online catalogues had to be studied. Thus, the subject of the present M.A. thesis was extremely topical in Estonia at the time when the author started to explore it.

This M.A. thesis aims to find practical models for subject description and subject searching in online catalogues, suitable for Estonian libraries. The work is based on international experience and our own ideas.

To this end, a number of tasks had to be done:

1. To study the essence of online catalogue as the core component of integrated library system and as an information retrieval system;
2. To analyse the elements and format of standard bibliographic records from the point of view of subject access;
3. To explore the types of controlled vocabulary and methods of verbal subject indexing that are suitable in electronic environment;
4. To determine the position of Universal Decimal Classification in the world, and its sustainability;

5. To identify international standards related to indexing and to analyse their applicability;
6. To study the possibilities for designing subject access and to evaluate optimal proportion of different retrieval methods.

The main result of the M.A. research was the designing of subject description and subject access model for the Estonian Libraries Network Consortium's shared databases. The basic principles of the model are the following:

1. Two mutually complementary methods – classification and subject indexing with descriptors – are to be used in subject description. Classification supports the systematic overview of the database's content and provides pre-coordinated terms for information search. Verbal descriptors provide an opportunity for free post-coordination of concepts in the search phase, both amongst themselves as well as in combination with keywords.
2. Common indexing languages are to be used in classification and subject indexing – the UDC Estonian edition and the Estonian Universal Thesaurus. To certain extent, free indexing terms are unavoidable.
3. Users have to be provided maximal support in searching for applicable access points, providing browsable indexes, integrated authority files, and instructions.
4. The authority control options provided by the system have to be used as widely as possible and libraries' classification and subject indexing practices have to be monitored. The Working Group of Classification and Subject Indexing involving all member libraries of the Consortium must have the central role in it.

As the objective of the research is directly related to library practice, its results have been implemented in many ways: in generating controlled vocabularies at Tartu University Library and the National Library of Estonia; in preparing the UDC Estonian edition; in designing subject access facilities for Tartu University's online catalogue INGRID, the Estonian Libraries Network Consortium's information system ESTER, information systems of public libraries, etc. Both the author as well as other

academic lecturers and educators have used the standpoints and publications, presented here, for the purposes of teaching.

The present M.A. thesis consists of two parts: “Libraries’ online catalogues” and “Indexing languages and subject access” An analytic overview is supplemented with publications, where the majority of the results of author’s research have been published.

PUBLIKATSIOONID

ESILEHELE
ENGLISH

OTSI

NUMBRID

infoFoorum

TPÜ R

ISSN 1406-9848

ARTIKLID

RAAMATUKOGUNDUS

Sirje Nilbe. Raamatukogu elektronkataloog 1
arengulugu ja olemus

Frederick C. Lynden. Ressursside jagamine: võimalus
üleminekuks paberkandjatelt elektroonilistele
(kokkuvõte)

INFOÜHISKOND

Voldemar Kolga. Infoühiskond ja toimetulek

Larissa Jõgi. Kuidas ja miks õppida armastama õppimist

Aivar Tõnso. Eesti ja infoühiskond

RAAMATUTEADUS

Mare Lott, Aile Möldre. Muutused Eesti
kirjastussüsteemis ja raamatutoodangus 1 1918-1940
(kokkuvõte)

English

Tagasi koduleheküljele

Kaastöö

Toimetus

Kontakt



Koostatud 06.06.2001
sfinx@tpu.ee

Koostatud 17 veebruaril 1997
sirvir@tpu.ee

RAAMATUKOGU ELEKTRONKATALOOG 1: ARENGULUGU JA OLEMUS

Sirje Nilbe

1. Sissejuhatus

Et tänapäeva muutuv maailmas elus püsida ja edasi areneda, on raamatukogud üha enam kanda võtnud infokeskuste rolli. Informatsiooni kogumine, töötlemine, säilitamine ja kättesaadavaks tegemine koosneb suurelt jaolt rutiinsetest toimingutest, mida arvutipõhise tehnoloogiaga saab tunduvalt ratsionaliseerida. Informatsiooni taasleidmiseks pakub kaasaegne infotehnoloogia kvalitatiivselt paremaid võimalusi kui varasemad "paberist" süsteemid.

Juba hulk aastaid on arvuteid kasutatud sellisteks tegevusteks nagu jadaväljaannete haldus ja registreerimine, laenutuste registreerimine, komplekteerimine, bibliokirjete koostamine, koondkataloogide pidamine jms. Meilegi on see tuttav asi. Lugejateni on aga raamatukogu sisetöö tulemused jõudnud ikka traditsioonilisel kujul: kartoteekide ja trükitud bibliograafiatena, trüki- ja kaartkataloogidena.

Lugeja infootsingu seisukohalt algas uus etapp raamatukogude automatiseerimises avalike siduskataloogide (*online public access catalogue/OPAC*) juurutamisega.

Käesolev kirjutis juhatab sisse artiklisarja, milles püütakse kirjanduse ja mõningase isikliku kogemuse põhjal anda ülevaade raamatukogu elektronkataloogide olemusest, arendamisest ja uurimisest. Eesti seni ainuke avalik siduskataloog töötab Tartu Ülikooli raamatukogus. See loodi täiesti oma jõududega ja on paljuski ebatüüpiline oma tehniliste lahenduste poolest, eriti võrreldes angloameerika malliga. 1997. aasta töötab tulla meie suurematele teadusraamatukogudele huvitav, sest alanud on ühise integreeritud raamatukogusüsteemi INNOPAC evitustöö. Vahest leiavad raamatukoguhoidjad neist kirjutistest natuke taustteadmisi, et valutumalt läbida keeruline periood üleminekul uuele töötehnoloogiale.

2. Elektronkataloogide tekkimine ja areng

Raamatukogude elektronkatalooge (EK) saab vaadelda kahest aspektist: kui infootsisüsteeme (IR-süsteemid) ja kui raamatukogutöö automatiseerimise põhikomponenti. Infootsisüsteemina peab EK täitma eelkõige neidsamu nõudeid, mis C. A. Cutter juba 1867. aastal raamatukogukataloogile esitas. Kataloog peab:

- võimaldama leida raamatut, millest on teada kas autor, pealkiri või teema;
- näitama, mis teoseid raamatukogus leidub teatud autorilt, teatud teemal või teatud laadi kirjandusest;
- abistama teose valikul kas bibliograafiliste, temaatiliste või kirjanduslike tunnuste järgi.

Raamatukogutöö tehnoloogia seisukohalt on EK andmebaas selleks tuumaks, millel rajaneb kogude haldus ja laenutus. EK-de sünnis oli määrava tähtsusega just viimane aspekt (vajadus ratsionaliseerida kataloogimist ja laenutuste kontrolli, raamatukogu igipõliseid tegevusi).

Kui kataloogima hakati arvuti abil, ei nähtud algul ette tekkiva andmebaasi kasutamist lugeja poolt. See oli mõeldud raamatukogu sisetöök, et hõlbustada traditsiooniliste kataloogivormide (kaart-, trüki- ja mikrovormkataloogide) tootmist. EK-de sünnimaal Ameerikas oli reeglilikohaste kirjade koostamine suurelt osalt koondunud kataloogimiskeskustesse, mis müüsid valmis kaarte või MARC-formaadis elektroonilisi kirjeid. Arvestuseks ja ringluse kontrolliks arvuti abil piisas raamatukogudel minimaalsete bibliograafiliste andmete sisestamisest teaviku kohta. Üsna varsti leiti, et need andmebaasid võiksid olla kättesaadavad ka lugejale nii raamatukogus kui sideliinide kaudu väljaspool raamatukogu. Elektronkatalooge hakati nimetama avalikeks siduskataloogideks (*online public access catalog - OPAC*).

Esimesed automatiseeritud süsteemid, mida võiks nimetada EK-deks, võeti kasutusele 1970. aastate keskel Ameerikas. Organisatsioonid, kus need juurutati, olid *Online Computer Library Center (OCLC)* (jaanuar 1974), *Research Libraries Group/Research Libraries Network (RLG/RLIN)* (detsember 1974) ja *Ohio State University Library* (jaanuar 1975) (Su 1994).

1970. aastate teine pool oli paljudes raamatukogudes katsetuste aeg. EK-de juurutamine ei läinud algul kuigi libedalt. Mõned raamatukogud loobusid EK-dest kui liiga kulukatest ja liiga "eksootilistest" süsteemidest. Osa raamatukoguhoidjaid pidas automatiseerimist tühjaks trikitamiseks. Kongressi Raamatukogu plaanis lõpetada kaartkataloogi täiendamise 1980. aastal, kuid valis jätkamiseks mitte EK, vaid arvuti abil toodetud mikrokaardid. Barjäär murdis 1980. aastate algul. 1982. aastal oli Ameerika Ühendriikides vähemalt 20 EK-d. Kui esimesed EK-d programmeeriti ainuüksnäpparidena kohalikuks kasutamiseks, siis 1985. aastal oli USA-s juba 48 firmat, kes müüsid kataloogisüsteeme (Su 1994). EK-d ja nende kasutamine lugeja poolt said ka teadusliku uurimistöö objektiks.

C. R. Hildreth, üks juhtivaid EK-de uurijaid, on juba 1983. aastast alates korduvalt rääkinud ja kirjutanud EK-de kolmest põlvkonnast. Tema jaotus põhineb EK-de kvalitatiivsetel erinevustel, sidumata ühtegi põlvkonda mingi kindla ajavahemikuga (Hildreth 1987). Hildrethi kriteeriumide järgi saab EK-

de arengu kokku võtta järgmiselt.

Esimesed EK-d tekkisid suurte raamatukogude laenutus- ja kataloogimissüsteemidest. Kuigi tasulisi infoteenuseid pakkuvaid infootsisüsteeme (*Information Retrieval system* - IR süsteem) oli käigus juba palju ja pikemat aega, ei võetud neid esialgu kuigivõrd eeskujuks EK-de loomisel. Peamine põhjus oli süsteemide erinev suunitlus: IR-andmebaasid olid algusest peale mõeldud infootsinguks, eriti aineotsinguks, EK-d aga kopeerisid kaartkatalooge. Erinevus nähtub juba bibliokirjete struktuurist: EK-s ei ole kokkuvõtteid ja sisuanalüüs on muuski osas palju napim kui kommertsandmebaasides. Sidusandmebaase vahendasid väljaõppinud infoteenindajad, kes said hakkama keerulise, kuid tulemusliku otsitehnikaga. Raamatukogu lugejalt ei saanud samasugust asjatundmist eeldada.

IR-andmebaase ja EK-sid võrreldes tuleb meele pidada ka seda, et raamatukogudel pole olnud nii palju raha arendustööks kui infofirmadel. Oluline vahe on seegi, et raamatukogus kataloogitav aine on palju mitmekesisem kui keskmise andmebaasi oma nii sisu, keelte kui ka teavikulaadide ja -tüüpide poolest.

Esimeste EK-de kirjed olid üsna napid: tihti vaid autor, pealkiri, kohaviit ja saadavusandmed. Märksõnu sageli polnudki. Ei olnud võimalik kasutada täistekstotsingut ega kärpimist (*truncation*), päring pidi täpselt vastama andmebaasi registri sõnadele või väljenditele. Järelkoordinatsiooni ei saanud üldiselt rakendada, mõnes süsteemis võis siiski otsida koos autori nime ja pealkirja algust.

Esimese põlvkonna EK-de loojad eeldasid, et kasutaja teab täpselt, mida otsib. Registrite või kirjete sirvimise (*browsing*) võimalust polnud või oli see väga piiratud. Otsiterminate valimiseks ega muukski süsteem abi ei pakkunud ja tihti peale lugeja ei saanudki aru, miks tema otsing ebaõnnestus.

Suhtlemine arvutiga käis enamasti menüürežiimis ja tulemuseks saadi lühikirjed, mis võimaldasid teose lokaliseerida.

Teise põlvkonna EK-des on eelmise põlve vigu püütud parandada, võttes eeskujuks IR-süsteemide omadusi. Kirjed on täielikumad ja otsitavaid välju on rohkem. On suurendatud registrite arvu ja täiustatud autoriteetfaile ning antud võimalus neid lehitseda. Aineotsing märksõnade ja võtmesõnade abil on tavaline ning täiendatud järelkoordinatsiooni ning Boole'i operaatoritega. Võtmesõnaregister on moodustatud tavaliselt pealkirja ja märksõnade väljadest. Otsitermineid saab paremalt poolt kärpida. Sageli on võimalik otsingut piirata ilmunisaja, ilmuniskoha, keele järgi.

Kasutajaliides on muutunud mitmekesisemaks ja interaktiivsemaks: juba registris antakse teada otsitunnusele vastavate kirjete arv, otsingu käigus on

võimalik tagasi liikuda, saab valida menüürežiimi (algajatele) ja käsürežiimi (edasijõudnutele) vahel jms. Arvutis on harjutusprogrammid ja eksimustest antakse teada koos parandusjuhistega. Leitud kirjed on võimalik saada mitmes formaadis (nt. lühiaandmed, täielik bibliokirje, MARC-kirje). Niisiis on teise põlvkonna EK-des ühendatud hästistruktureeritud raamatukogukataloogi ja konventsionaalse Boole'i otsingul põhineva IR-süsteemi omadused. Enamik praegu töötavaid EK-sid on seda tüüpi.

Kui lisada teise põlve omadustele mitmeid täiendusi, peaksid EK-d muutuma veel efektiivsemaks ja kasutajasõbralikumaks ning moodustama kolmanda põlvkonna. Oluliseks peetakse, et päringuid saaks esitada loomulikus keeles, et andmebaasis navigeerimist assisteeriks intelligentsed programmid, et otsingu igas järgus oleks hea tagasiside. Aineotsingus näiteks tuleb integreerida märksõna-, liigi- ja võtmesõnaotsing, nii et otsiterminite konversioon toimiks automaatselt. "Täpse vaste" (*exact-match*) printsiibi asemel peaks süsteem toimima "lähima vaste" (*best-match*) põhimõtte järgi. Väljundkirjed tuleks reastada vastavalt nende oletatavale relevantsusele ja kasutajal peaks olema võimalus anda tagasisidet väljundi relevantsuse kohta (*relevance feedback*).

Kataloog peaks hõlmama raamatukogu varasid täielikult, osalt bibliokirjetena, osalt täistekstidena ja kujutistena. Bibliokirjed võiks olla inforikkamad. EK-le peaks olema laialdane juurdepääs infovõrkude kaudu, teisalt peaks kataloog olema seotud mitmesuguste teiste andmebaaside ja infosüsteemidega.

Kolmanda põlvkonna EK-sid, millel oleks kasvõi enamik loetletud omadusi, praegu veel ei eksisteeri. Prototüüpsüsteeme, milles katsetatakse üht või mitut uuendust, on küll loodud, kuid raamatukogusüsteeme tootvad firmad on üsna visad oma toodangut kardinaalselt muutma, olemata kindlad, et see nende läbimüüki oluliselt suurendaks. Integreeritud raamatukogusüsteem on üks keerukamaid infosüsteeme, ja raamatukogud mitte just eriti maksujõulised kliendid.

Raamatukogu seisukohalt vaadates leiab samuti mitmeid takistusi EK-de põhimõtteliseks uuendamiseks. Kataloog on kumulatiivne andmebaas, mille kirjete, sisuanalüüsi ja autoriteetfailide tagantjärele täiendamine ei tule suuremas mahus kõne alla, eriti kui järjekorras ootamas hulk konverteerimata materjali ja uudiskirjandust. Kui on vaja valida, kas eelistada sisetöö ja bibliograafilise kontrolli või lugeja vajadusi, kaldub kogu maailma raamatukoguhoidjate südametunnistus esimese alternatiivi poole, kasvõi juba seepärast, et oma vajadused on paremini ja ühemõttelisemalt teada.

Nimetatud vastuolud ei tähenda aga, et EK-de areng oleks lõplikult seiskunud. Infootsing kui uurimisala areneb jõudsalt ning küllap jõuavad tulemused ka raamatukogukataloogidesse. Meie oludes oleks suur edasimineku teadusraamatukogudes korralikult käima panna teise põlvkonna omadustega

elektronkataloog ja see võimalikult kiiresti rohke infoga täita.

3. Elektronkataloogi kontseptuaalne mudel

Iga EK-i, hoolimata selle konkreetsetest iseärasustest, saab vaadelda kui funktsionaalsete kihtide (*layers*) süsteemi, mis vahendab kasutajat ja andmebaasis säilitatavaid bibliograafilisi andmeid. Need funktsionaalsed kihid on järgmised (Larson 1991):

- kasutajaliidese kiht (*user interface layer*), mis võimaldab lugejal suhelda EK-i ülejäänud osadega;
- andmebaasihaldussüsteemi liidese kiht (*database management system interface layer*), mis tagab kasutajaliidese ja andmebaasihaldussüsteemi vahelise suhtlemise;
- andmebaasihaldussüsteemi kiht (*database management system layer*), mis tegeleb andmebaasi salvestatud informatsiooni korrastamise, otsingu ja leidmisega;
- andmebaasikiht (*database layer*), mis säilitab bibliograafilist ja abiinformatsiooni.

Funktsionaalsed kihid on realiseeritud infootsisüsteemina vastava tarkvara ja riistvara abil. Realisatsiooni tasandil on kihtide ülesanded lahendatud eri süsteemides erinevalt ja kihid ise pole nii selgelt eristatavad.

Raamatukogukataloogile omast informatsiooni sisaldab ja säilitab niisiis ainult andmebaasikiht, kuna ülejäänud kihtide ülesanne on seda informatsiooni korrastada ja kasutajale väljastada. Funktsionaalsete kihtide, eriti andmebaasi, suhtelise iseseisvuse tõttu on võimalik näiteks kirjeid ühest süsteemist teise üle kanda.

EK andmebaasis säilitatavad infokogumid on tavaliselt kolme tüüpi: bibliokirjed, registrid ja autoriteetfailid. Lähemalt nendest edaspidi.

Andmebaasihaldussüsteemi (*Database Management System/DBMS*) kihi ülesandeks on:

- uute kirjete lisamine andmebaasi ja nendes kirjetes leiduvate andmete indekseerimine (registrite täiendamine);
- kirjete kustutamine andmebaasist ja registrite muutmine vastavalt sellele;
- kirjete parandamine ja vastavate registrite muutmine;
- kirjete otsimine ja leidmine andmebaasist registrite abil otsitunnustele antud väärtuste järgi.

Lugejad saavad kasutada vaid EK otsifunktsiooni. Kirjete lisamine, parandamine ja kustutamine on võimalik ainult selleks volitatud

personaliliikmetel. Andmebaasi hooldus toimub sõltuvalt tarkvarast kas reaalajas, mispuhul muudatused andmebaasis ilmnevad vahetult pärast nende tegemist, või pakktöötlusena (*batch processing*), mil sisestatud muudatused lisatakse andmebaasi portsukaupa näiteks öhtul või öösel; süsteemi otsifunktsioon on sel ajal blokeeritud.

Andmebaasi struktuur ja otsimehhanismid võivad olla küllalt keerulised, kuid kasutaja ei pea neid tundma.

Andmebaasihaldussüsteemi liidese kihi ülesanne on tõlkida kasutaja käsud või menüüvalikud vastavaks reaks andmebaasioperatsioonideks. See rida võib sisaldada mitmete pöördus- ja otsimeetodite, registrite ja andmebaasifailide koordineeritud kasutamist selleks, et anda vastus lugeja päringule.

DBMS-i liidese kihi ülesandeks on ka sellised otsinguga seotud toimingud, mida ei soorita DBMS ise, näiteks loogikaoperatsioonid eelnevalt väljaotsitud kirjade hulkadega või kirjade reastamine relevantsuse järjekorras.

DBMS-i liides võib osaliselt või tervikuna paikneda isegi hoopis teises arvutis. Sellisel juhul peab liides sisaldama programme, mis konverteerivad päringud ja vastused võrguülekandeks vajalikku standardvormingusse.

Kasutajaliidese kihti kuuluvad need süsteemi osad, mis vahetult tegelevad infovahetusega kasutaja ja süsteemi vahel. See on EK kõige nähtavam osa ja mõjutab, nagu näitavad uuringud, kõige enam tavakasutaja arvamust kataloogist, kuigi andmebaasi sisu ja süsteemi otsiomadused on tulemuste seisukohalt olulisemad.

Kasutajaliidese põhiülesanded on:

- kasutaja käskude vastuvõtt ja analüüs. Käsud võivad olla antud formaalses sisendkeeles, menüüvalikutena, funktsiooniklahvidega, kursori ja hiireklõpsu abil või nende kombinatsioonidena;
- tuvastada, milliseid tegevusi lugeja nende käskudega süsteemilt soovib ja stimuleerida DBMS-i liidest ja DBMS-i sooritama vastavaid protseduure;
- korraldada ja vormindada väljaotsitud bibliograafilised andmed ja süsteemi teated lugejale väljastamiseks ja esitada need kuvari ekraanil;
- leida üles vead kasutaja käskudes ja anda abi korrektsete käskude vormistamiseks.

Hästi projekteeritud ja kergesti käideldav kasutajaliides võib varjata otsimehhanismide sisemise keerukuse või korvata nende mehhanismide puudujääke inimese ja arvuti interaktsiooni parandamise teel.

Esitatud kontseptuaalne mudel peaks aitama arvutihariduseta

raamatukoguhoidjatel, kes kataloogi koostavad, kasutavad, lugejatele vahendavad ja täiustusi planeerivad, mõista, mis põhimõtteliselt toimub selles "mustas kastis", millest pole näha muud kui see, mis kuvarile ilmub.

Kasutatud kirjandus

Hildreth, C. (1987). Beyond Boolean: designing the next generation of online catalogs. *Library Trends* 45 (4) 647-667

Larson, R. R. (1991). Between Scylla and Charybdis: subject searching in the online catalog. *Advances in Librarianship* 15 175-236.

Su, S.-F (1994). Dialogue with an OPAC: how visionary was Swanson in 1964? *Library Quarterly* 64 (2) : 130-161

English

Tagasi koduleheküljele

Koostatud 17 veebruaril 1997
sirvir@tpu.ee

ONLINE PUBLIC ACCESS CATALOGUES 1: NATURE AND DEVELOPMENTS

Sirje Nilbe

The paper introduces a series of review articles about OPACs: what they are, how do they function, which problems do they cause for information retrieval, what are the main directions of OPAC research and development.

The first part of the paper deals with origin and evolution of online catalogues from simple circulation or cataloguing systems to complicated information retrieval systems. Three generations of online catalogues according to C. Hildreth's conception are described. Today most of the operating OPACs have only the features of the second generation but there is a number of prototype systems with enhanced retrieval capabilities in many libraries.

In the second part of the paper the conceptual model of an OPAC is introduced as it is presented by R. R. Larson. By Larson, every online catalogue system can be seen as a set of functional layers mediating access between the user and the bibliographic data stored in the database.

These functional layers are: the user interface layer, the DBMS interface layer, the DBMS layer, and the database layer.

Estonian

[Back to the homepage](#)

Last updated February 17. 1997
sirvir@tpu.ee

ESILEHELE
ENGLISH

infoFoorum

TPÜ

ISSN 1406-9848

OTSI

NUMBRID

ARTIKLID

RAAMATUKOGUNDUS

Sirje Nilbe. Raamatukogu elektronkataloog 2:
arengulugu ja olemus

Silvi Roomets, Krista Talvi. Raamatukoguhoidjate
täienduskoolitus Eestis

INFOTEADUS

Keiu Saarniit. Uued võimalused meditsiiniinfo
valdkonnas Tartu Ülikooli Kliinikus

RAAMATUTEADUS

Mare Lott, Aile Möldre. Muutused Eesti
kirjastussüsteemis ja raamatutoodangus 2: 1918-1940
(kokkuvõte)

Ausra Navickiene. Raamatukaubandus Leedus
üheteistkümnenda sajandi esimesel poolel
(kokkuvõte)

English

Tagasi koduleheküljele

Koostatud 31. juulil 1998
sirvir@tpu.ee

Kaastöö

Toimetus

Kontakt



Koostatud 06.06.2001
sfinx@tpu.ee

RAAMATUKOGU ELEKTRONKATALOOG 2: ANDMEBAAS

Sirje Nilbe, Eesti Rahvusraamatukogu normandmete osakonna juhataja

"Infofoorumi" teises numbris oli juttu elektronkataloogide (EK-de) kujunemisest ja funktsionaalsest struktuurist. Seekord vaadelgem, milliseid andmeid säilitatakse EK andmebaasis ja kuidas neid vormistatakse.

Tavapäraseimad andmekogumid EK-s on bibliikirjed, autoriteetfailid ja registrid.

Bibliikirjed

Kõige tähtsamad andmed, mida elektronkataloogi andmebaasis hoitakse, on loomulikult bibliikirjed. Bibliikirjete põhisisu pole arvutiajastul muutunud võrreldes kaartkataloogide ja trükibibliograafiatega. Kirjed koostatakse ikka kirje- ja kataloogimisreeglite järgi, kuid andmete arvutitöötlus ja infovahetus raamatukogude vahel seab suuremad nõuded kirjete korrektsusele ja reeglite ühtlustamisele.

Tänapäeval on peaaegu kõigis maades kataloogimisreeglite aluseks ISBD-d (*International Standard Bibliographic Description*). ISBD-d on IFLA poolt välja töötatud standardite rühm, mis määrab eri teavikutüüpide kirjeldamisel esitatavate andmete valiku, vormi ja järjekorra, samuti kirjavahemärgid, mida kirjes kasutatakse. ISBD-d ei käsitle pealdisi, sisuavamiselemente ega muud täiendavat teavet, mis kataloogikirjes antakse. Teisalt on eri maade kataloogimisreegleid *de facto* standardina mõjutanud angloameerika reeglid (AACR, *Anglo-American Cataloguing Rules*). Nende teine, uuendatud väljaanne AACR2 põhineb kirje osas samuti ISBD-le, kuid sisaldab lisaks pealdiste, s.t. otsitunnuste moodustamise reegleid. Seega on praeguseks maailmas saavutatud märkimisväärne ühtsus teavikute kirjeldamisel ja otsielementide määramisel.

Kirjete juurde kuuluv sisuanalüüs pole kaugeltki nii hästi reeglistatud. Teavikute sisu avamises valitseb maailmas suur kirevus, selle standardiseerimine on raske, kui mitte võimatu ülesanne. Kõige üldisemalt jaguneb intellektuaalne (inimese tehtav) sisuanalüüs liigituseks ja märksõnastuseks, mille tulemusel lisatakse bibliikirje kirjeldavale osale liigiindeksid ja märksõnad. Detailsemat sisututvustust, nt. annotatsioone, sisukokkuvõtteid, sisukordi, kataloogikirjest tavaliselt ei leia. Nende lisamine tuleb kõne alla vaid laia teeninduspiirkonnaga keskkataloogimisasutuste olemasolu puhul, kuid igal juhul aeglustab see infotöötlusprotsessi tunduvalt.

Ühtsed kirjereeglid on loodud silmas pidades bibliikirjete levitamist

paber kandjal. Kirjete arvutitöötlamine esitab nende vormile ja struktuurile oma nõuded, mis johtuvad osalt arvuti tööpõhimõtetest, osalt neist funktsioonidest, mis masinloetaval bibliokirjel täita tuleb.

Kaartkataloogimise ja arvutikataloogimise alused on mitmes suhtes erinevad. Kaartkataloogimise eesmärgiks on "valmistoote" - kataloogikaardi koostamine, arvutikataloogimisel ainult määratletakse bibliograafilised andmed ja varustatakse need tunnuskoodidega. Iga teaviku kohta on vaja koostada vaid üks põhikirje, millest arvuti abil võib saada mitmesuguseid "valmistooteid" - erineva pealdisega kataloogikaarte, bibliograafiaid, nimestikke, inventariraamatuid. Põhikirjeid saab kasutada siduskataloogi andmebaasis, laenutuste kontrollis jms. Koopiaid levitamiseks saab teha kiiresti ja kvaliteetselt.

Masinloetav kirje peab olema arvutisüsteemile täiesti ühemõtteliselt esitatud. Kaartkataloogi kasutaja mõistab paljusid andmeid oma eelteadmiste põhjal. Näiteks on teose keel selline teave, mida kaardil ei ole eraldi mainitud. Kirje alad ja elemendid tuntakse ära nende sisu, asukoha ja järjestuse järgi kaardil. Arvutisüsteemi jaoks tuleb kõik andmed täpselt kodeerida, et see oskaks neid õigesti interpreteerida. Peale selle poleks masinloetavate kirjete koostamisel erilist mõtet, kui neid ei saaks magnetkandjal või sideliinide kaudu levitada. Seepärast kehtib juba 1981. aastast ISO standard "*Format for bibliographic information interchange on magnetic tape*" (ISO 2709), mis on kõigi arvutipõhiste kataloogimissüsteemide nurgakivi.

ISO 2709 määrab kindlaks kirje põhistruktuuri ja mitmesuguste tehniliste andmete esitusviisi. Kirje täpsem struktuur ja kasutatavad koodid määratletakse nimetatud standardi rakendustes - bibliograafilistes formaatides, millest tuntuimad kuuluvad MARC-rühma.

Algne MARC (*Machine Readable Cataloguing*) loodi aastail 1965-1966 Kongressi Raamatukogus, kus katsetati kataloogikirjete tootmist arvuti abil. Suurbritannias tegeldi samalaadse projektiga, eesmärgiks kasutada masinloetavaid kirjeid Briti rahvusbibliograafia trükiväljaande tegemisel. 1968. aastal algas koostöö, mille tulemusena loodi nn. MARC II. Selle funktsioon pidi olema eelkõige kommunikatiivne - MARC oli defineeritud kui teabeedastusformaat. Hiljem Ameerika ja Briti MARC-id siiski lahknesid ja nii tekkisid USMARC ja UKMARC. Nende eeskujul on paljudes maades loodud oma rahvuslikud formaadid arvestades kohalikku kataloogimispraktikat ja muid vajadusi (nt. CanMARC, FINMARC, Saksamaal MAB).

Rahvusvahelist infovahetust selline paljusus muidugi takistab. Võõraid MARC-e tuleb tundma õppida, kirjete ülevõtmiseks tuleb kirjutada konverteerimisprogrammid, mis pole alati lihtne. Viimastel aastatel toimib taas ühtlustumistendents, sest bibliokirjete rahvusvaheline ühiskasutus üha

levib. Euroopa ühisformaadiks näib kujunevat IFLA poolt hallatav UNIMARC, mida soosib ka Euroopa Liit. Paljud Kesk- ja Ida-Euroopa maad pole hakanud oma rahvuslikke formaate välja arendama, vaid on võtnud kasutusele UNIMARC-i või USMARC-i. Mitmed Lääne-Euroopa riigidki on üle läinud UNIMARC-ile. USMARC-i CanMARC-i ja UKMARC-i integreerimiseks on käimas pikaajaline programm.

Eestis võetakse teadusraamatukogudes kasutusele USMARC, sest sellel põhineb ostetav raamatukogusüsteem INNOPAC. TÜ raamatukogu praegune elektronkataloog toetab FINMARC-i, nii et konverteerimisprobleemid on meilgi kohe omast käest võtta.

Järgnevalt on esitatud üks näide USMARC-formaadis kirje kohta, mis traditsioonilise põhikirjena näeks välja nii:

Saaksin ma saksa sundijaks / Eesti NSV Teaduste Akadeemia Keele ja Kirjanduse Instituut ; toimetanud Eduard Laugaste. - Tallinn : Eesti Raamat, 1976. - 427 lk. : ill. ; 23 cm. - Kokkuvõtted vene ja saksa keeles. - Bibliograafia lk. 417-425. - (Uurimusi eesti regivärsi ja rahvajutu alalt ; 1).

MARC-kirje:

LDR *****nam##22*****5i#*****

DIR (süsteemi poolt genereeritud kataloog)

001 (kontrollnumber)

003 (kontrollnumbri identifikaator)

005 (viimase tehingu kuupäev ja kellaaeg)

007 ta

008 970401 s1976#### ee#a#### ##### #000#0# est#d

041 0# SaestSbgerrus

080 ## \$a398.8.09(=511.113)(082)

245 00 \$aSaaksin ma saksa sundijaks /\$cEesti NSV Teaduste Akadeemia Keele ja Kirjanduse Instituut ;. toimetanud Eduard Laugaste

260 ## \$aTallinn :\$bEesti Raamat,\$c1976

300 ## \$a427 lk. :\$bill. ;\$c23 cm

440 #0 \$aUrimusi eesti regivärsi ja rahvajutu alalt ;\$v1

504 ## \$aBibliograafia lk. 417-425

520 ## \$aKokkuvõtted vene ja saksa keeles

650 #4 \$aorjuslaulud

650 #4 \$aregivärsilised rahvalaulud

651 #4 \$aEesti

655 #7 \$aartiklikogumikud

700 1# \$aLaugaste, Eduard,\$etoimetaja

710 2# \$aEesti Keele Instituut

Kirje esimene pool sisaldab mitmesuguseid kodeeritud andmeid. Paljud neist genereeritakse süsteemi poolt tehniliseks otstarbeks, aga osa iseloomustavad ka kataloogitavat infokandjat. Näiteks kood *a* kirje päises (*leader* lüh. LDR) 7 kohal näitab, et tegemist on keelelise materjaliga, sellele järgnev *m* tähendab monograafilist kirjeldustaset. Väli 007 on ette nähtud materjali füüsilise kirjelduse jaoks ja sellel asuvad koodid *t* ja *a* tähistavad teksti harilikus trükikirjas. Välja 008 kuus esimest kohta on mõeldud kirje koostamise kuupäeva jaoks, järgmise koodirühmaga tähistatakse trükise ilmumisaastaid. Järgmise rühma alguses on ilmumismaa kood *ee* (Eesti), järgnev kood *a* näitab, et trükises leidub illustratsioone jne. Viimases rühmas olev kood *est* tähistab teose põhikeelt. Märkiga # on siin näites tähistatud tühikud, s.o. formaadis defineerimata positsioonid ja defineeritud, kuid antud juhul väärtust "tühik" omavad positsioonid.

Edasi tulevad väljad, mis koosnevad välja märgendist ehk numbrist, kahest indikaatoripositsioonist, mis vajaduse korral täpsustavad järgnevate andmete iseloomu või päritolu, ja alamväljadeks jaotatud andmetest. Alamvälja algust tähistab näites märk \$, millele järgneb üks araabia väiketäht. Väljal 041 on teose keeled kodeeritud kujul: alamväljal *a* eesti keel põhikeelena ja alamväljal *b* saksa ja vene keel kokkuvõtete keelena. 080 on UDK indeksi jaoks. Traditsioonilised kirjeandmed algavad väljaga 1XX, kuhu pannakse põhipealdis. Antud näites põhipealdis ei ole, sest kirjereeglite järgi tuleb trükist kirjeldada pealkirja järgi.

245 on pealkirja ja vastutusandmete väli. Mõlemad indikaatorid on siin väärtusega 0, mis tähendab, et pealkirjale ei tehta täiendkirjet ja järjestamist

alustatakse pealkirja esimesest tähest. Pealkiri on alamväljal *a* ja vastutusandmed alamväljal *c*. Kirjavahemärkidest niipalju, et kirjeala-sisesed märgid tuleb tavaliselt kirjutada käsitsi, kuigi mõni süsteem võib panna need ka automaatselt. Väljal 260 on ilmunisandmed, väljal 300 arvandmed. Väli 440 on siin seeriaandmete jaoks, toimides ühtlasi sarja täiendkirje pealdisena. Väljad 5XX on mitmesuguste märkuste jaoks ja erineva tähistusega selleks, et neist paremini aru saada ja vajaduse korral erinevalt töödelda. Osa märkusi näiteks võib olla kasulik indekseerida ja seega nendes esinevad sõnad otsitavateks teha, mõnda aga vahest ei tahetagi avalikus kirjes näidata, vaid hoida "ametialaseks kasutamiseks"

Väljad 6XX on märksõnade jaoks. Märksõnadena on USMARC-is tõlgendatud kõiki verbaalseid väljendeid, mis on kirjele lisatud tähistamiseks teose sisu ja vormi - seega ka nimesid ja pealkirju käsitlese ainenäna. Antud näites on teosele omistatud kaks temaatilist märksõna (väljad 650), kohanimi märksõnana (väli 651) ja vormimärksõna (väli 655). Teise indikaatori abil saab ära näidata märksõna allika; väärtus 4 tähendab, et allikas (märksõnastik) ei ole määratletud.

Kirje lõpus on kaks täiendkirje pealdist, väljal 700 toimetajale ja väljal 710 kollektiivautorile. Kollektiivi nimi on esitatud kirje koostamise ajal kehtinud vormis. Indikaator 1 isikunime ees märgib, et nimi algab üheosalise perekonnanimega, indikaator 2 kollektiivinime ees seda, et sõnad on pärijärjestuses, s.t. pole kasutatud inversiooni.

Sellise kirje koostamine ja lugemine tundub üsna keeruline ja tülikas, aga tuleb harjuda. Pealegi on enamik kataloogimissüsteeme seadistatud nii, et osa koodide väärtusi antakse vaikimisi ette ja ka kõiki märgendeid pole vaja endal kirjutada. Korduvaid kirjeelemente nagu nimed, jadaväljaannete pealkirjad, märksõnad, liigiindeksid, saab oma kohale "tõsta" autoriteetfailist. MARC-vorming on vaid üks võimalikest kirje ekraanile kuvamise viisidest, mida lugejakataloogis ja enamikul raamatukoguhoidjatest pole õigupoolest vajagi. Ainult töötledjad peavad MARC-i hästi tundma, vähemalt seda osa, kuhu lähevad nende sisestatud andmed. Aga üldise ettekujutuse sellest, mis MARC-kirje on, milliseid andmeid sisaldab ja kuidas nende abil kataloogiotsingut teha, peaks küll iga raamatukoguhoidja endale looma.

Autoriteetfailid

Autoriteetfailid luuakse selleks, et hõlbustada autoriseeritud andmete ohjet ja kasutamist kataloogimisel. Autoriseerimist vajavad need kirje andmed, mis on otsitavad (*access points*) pealdistena: isikunimed, kollektiivnimed, ühtluspealkirjad, märksõnad, tihti ka liigiindeksid. Pealdise autoriseeritud kuju abil koondatakse kokku kõik antud autori või kollektiiv autori teosed, samuti nt. klassikalise teose eri väljaanded, vaatamata sellele, millist kuju on kasutatud konkreetsetes teoses. Variantvormid hoitakse autoriteetfailis

äraviidetenä. Märksõnade ja liigiindeksite autoriseerimisega saab parandada sisuanalüüsi järjepidevust. Seda tööd on tehtud ka kaartkataloogimise ajal, kuid arvuti pakub siingi palju rohkem võimalusi kui abikartoteegid.

Autoriseeritud andmed on soovitatav vormindada nii, et neid saaks kasutada bibliokirjete koostamiseks ja infootsinguks. Selleks on loodud omaette formaadid, mis ühilduvad vastava bibliokirje MARC-formaadiga, nt. *USMARC Format for Authority Data* ja *UNIMARC/Authorities*. Kui ühes ja samas andmebaasis on ühilduvates formaatides bibliokirjed ja autoriteetandmete kirjed, ei pruugi bibliokirje pealdist füüsiliselt sisaldadagi, vaid piisab autoriteetkirje kontrollnumbrist, mille abil pealdis bibliokirje vaatamise ajal nähtavaks saab.

Järgnev näide kujutab endast USMARC-formaadis autoriteetkirjet kollektiivinime *Eesti Keele Instituut* kohta, mida eelnenud bibliokirjenäites kasutati täiendkirje pealdisena.

LDR ***nz###22*****0###*******

DIR (süsteemi poolt genereeritud kataloog)

001 (kontrollnumber)

003 (kontrollnumbri identifikaator)

005 (viimase tehingu kuupäev ja kellaaeg)

008 970401 #n#az znnaa bn### ##### ##fa# ana## ##bd

110 2# \$aEesti Keele Instituut

410 2# \$aKeele ja Kirjanduse Instituut

410 2# \$aEesti NSV Teaduste Akadeemia.\$bKeele ja Kirjanduse Instituut

410 2# \$aEesti Teaduste Akadeemia.\$bKeele ja Kirjanduse Instituut

410 2# \$aEesti Teaduste Akadeemia.\$bEesti Keele Instituut

Nagu näha, sarnaneb autoriteetkirje põhistruktuur bibliokirje põhistruktuuriga. Kood *z* päises 7 kohal näitab, et tegemist on autoriteetkirjega, number 0 18. kohal tähistab lõpetamata kirjet. Välja 008 alguses on kirje tekitamise kuupäev. Kood *a* teises koodirühmas tähendab kinnitatud pealdist. *a* ja *a* kolmandas grupis näitavad, et väljendit võib kasutada nii deskriptiivse pealdisena kui ka märksõnana jne.

Väljal 110 on kollektiivinime autoriseeritud kuju, järgnevatel väljadel 410 selle variantvormid, mida kataloogimisel pealdiste ega märksõnadena ei kasutata, s.t. sisuliselt äraviited. Autoriteetkirje struktuuris on kohad ka definitsioonidele, juhendavatele ja muudele märkustele, aga siin pole neid kirja pandud.

Andmete autoriseerimine on küllalt töömahukas asi, kuid korralik autoriteetfail kergendab tunduvalt kataloogijate tööd ja parandab kataloogi kvaliteeti.

Registrid

Kolmanda andmekogumi, mida EK andmebaasis hoitakse, moodustavad registrid (*indexes*). Registrid tekitatakse bibliokirjete indekseerimise teel. See tähendab, et andmebaasihaldussüsteem koostab etteantud parameetrite järgi olulisematest kirjeandmetest loendid, mille abil ta hiljem leiab üles vajalikud kirjed. Registrite eesmärk on kiirendada otsingut, ilma nendeta peaks süsteem mingi tunnuse otsimisel läbi vaatama kõik kirjed, aga see võtaks vähegi suuremas andmebaasis liiga palju aega.

Kõige tavalisem registri vorm on nn. pöördfail (*inverted file*). Pöördfail koosneb otsitunnustest ja viitadest, mis juhatavad antud tunnuseid sisaldavate kirjete juurde. Viitadeks võivad olla kirjete identnumbrid või aadressid, mis näitavad kirje asukohta andmebaasifailis. Seega sarnanevad EK registrid trükistes, nt. raamatutes või bibliograafiates leiduvate registritega: raamaturegistris on aadressiks lehekülje number, bibliograafianimestikes kasutatakse viitadena tavaliselt kirjenumbreid. Vahe on selles, et EK registrid moodustatakse ja uuendatakse automaatselt.

Levinuimad registrid EK-des on autoriregister, pealkirjaregister, märksõnaregister, liigiregister ja võtmesõnaregister. Loendeid võib moodustada ka teistest kirjeelementidest. Kõik sõltub sellest, milliseid otsitunnuseid peetakse antud süsteemi võimaluste piires otstarbekaks kasutajale esitada.

Samanimeliste registrite sisu võib süsteemiti erineda olenevalt sellest, millised MARC-formaadi väljad on indekseerimiseks valitud. Näiteks võib kollektiivautorid paigutada koos isikunimedega autoriregistrisse või moodustada neist omaette loendi. Isikunimed käsitlekse ainena (personaalia) on mõnes kataloogis märksõnaregistris, teises jälle on moodustatud ühtne isikunimeregister, mille kaudu saab korraga kätte nii autori poolt loodud kui ka tema kohta kirjutatud teosed. Pealkirjaregister võib sisaldada nii üksikteoste kui ka jadaväljaannete pealkirju, teisel võivad viimased olla eraldi loendiks.

Otsivõimalused, mida lugejale pakuvad autori-, pealkirja-, märksõna- ja

liigiregistrid, on olemas või vähemalt teostatavad ka kaartkataloogis. Elektronkataloogis täiendab neid (või on kasutussageduselt tihti lausa esikohal) võtmesõnaregister, mis moodustatakse valitud kirjeväljadel asetsevate keeleliste ühendite üksikutest sõnadest. Kõige kasulikumad otsinguks on muidugi igasuguste pealkirjade sõnad, aga ka kollektiivinimedes, märkustes jms. sisalduvad sõnad.

Eelpool esitatud kirjenäite teose ja sarja pealkiri annaks võtmesõnaregistrisse järgmised sõnad (ase-, side- ja kaassõnu ning muid otsiväärtuseta sõnu ei indekseerita; nende vältimiseks antakse süsteemile ette vahelejäätuloend):

eesti

rahvajutu

regivärsi

saaksin

saksa

sundijaks

uurimusi

Sellest näitest ilmneb ka võtmesõnaindekseerimise üks puudusi - see võib tekitada müra ja valseoseid. Kogumiku pealkirja sõnad *saaksin*, *saksa*, *sundijaks* on võrdlemisi ebaadekvaatsed aineotsingu jaoks. Võtmesõnastamise tulu sõltub ka keelest - eestikeelse teksti "hakkimine" viib loendisse ühest ja samast sõnatüvest palju erinevaid vorme, tuletisi ning liitsõnu eraldi sissekannetena, mida ei saa ilma lisaabinõudeta koos otsida.

Registrid on vajalikud nii andmebaasihaldussüsteemile kirjete ülesleidmiseks kui ka kataloogi kasutajale oma päringu väljendamiseks otsiterminite abil. Kasutajale kuvatud registrivormingus pole viitasid näha, neid on vaja vaid otsimisprogrammile. Seevastu tunnusele vastavate kirjete üldarv pakub küll huvi kui eelorientiir otsingu tulemuslikkuse hindamiseks ning see arv on kasutajakuval tavaliselt olemas.

Elektronkataloogi andmebaasi sisu on seega rikkalikum ja keerukam kui esmapilgul, vaid lugejakataloogi moodulit kasutades paistab. Keerukus pole tingitud mitte niivõrd arvutitöötamise nõuetest, kuivõrd kataloogi andmebaasis hoitava informatsiooni iseloomust ja mitmekülgse taasleidmise vajadusest.

English

Tagasi koduleheküljele

Koostatud 31 juulil 1998
sirvir@tpu.ee

ONLINE PUBLIC ACCESS CATALOGUES 2: THE DATABASE

Sirje Nilbe, National Library of Estonia

Abstract

The core of an OPAC is the database which usually includes data of three types - the bibliographic records, the authority records, and the indexes.

The bibliographic records are created according to the cataloguing rules like ISBD and AACR2, and formatted for computer processing according to the particular bibliographic format like USMARC or UNIMARC. The authority records serve for authority control by cataloguing and searching. They are created according to the MARC formats for authority data compatible with corresponding bibliographic formats. The general form for indexes in a bibliographic database is an inverted file including the access points to the records and the pointers referring to the relevant records.

In the paper an introduction to the MARC, authority control and computer-based database indexing is provided.

Estonian

[Back to the homepage](#)

Last updated July 31, 1998
sirvir@tpu.ee



1/1995



Kataloogimise automatiseerimisest Tartu Ülikooli raamatukogus;

Mis toimub Tartu Ülikooli raamatukogu hoidlates?;

Milleks veel "Mana", kui "Tulimuld" olemas? jm.

Avaveerg	Ivi Eenmaa	4
TEADUSRAAMATUKOGU		
Et elule mitte jalgu jääda	Urve Tõnnov	5
Kataloogimise automatiseerimisest		
Tartu Ülikooli raamatukogus	Elsa Loorits, Sirje Nilbe	9
Dokumentide sisu avamine ja aineotsing arvutikataloogis	Sirje Nilbe	12
Mis toimub TÜ raamatukogu hoidlates?	Urmas Kokassaar, Kurmo Konsa	14
Eesti Riigikaitse Akadeemia raamatukogu	Sigrid Mandre	17
Eesti Kirjandusmuuseumi Arhiivraamatukogu raamatuid	Mihkel Volt	18
KULTUURISUHTED		
Leedu rahvaraamatukogude tänapäev	Vitautas Gudaitis	19
JUTUTUBA	Maire Liivamets	21
ARVUSTUS		
Milleks veel "Mana", kui "Tulimuld" olemas?	Maie Kalda	22
Eesti õigusbibliograafia nimestikku sirvides	Arvo Junti	24
TEIST JA TEILE		
Kukeristi Kaalu ja kirjandusuurija	Jaanus Vaiksoo	25
Lugeja arvab:		
Kutseksam on raamatukogus töötamise garant	Mati Muru	26
JURISTIVEERG		
Ülevaade töölepingu seadusest	Heli Naeris	27
KOMMENTAAR		
Kooliajakirjadest	Hans Jürman	28
Täienduseks & täpsustuseks	Olev Esna	28
MÄLESTAME		
Kalju Oja	Hans Jürman	30
Leida Püss	Mari Sibul	31
SÕNUMID		
Asutati Eesti Hoiuraamatukogu		31
Eesti kooliraamatukogud 1994. aastal		36
TASULINE LEHEKÜLG		
Tartu Ülikooli Raamatukogus koostatavad bibliograafiakartoteegid		37
SUMMARY		38
PE3IOME		39

KATALOOGIMISE AUTOMATISEERIMISEST TARTU ÜLIKOOLI RAAMATUKOGUS

ELSA LOORITS, SIRJE NILBE

TÜR

Kataloogimist automatiseeritakse, et kokku hoida raamatukogu tööjõudu ja tarbija infootsinguteks kuluvat aega, ning see on üks osa raamatukogu komplekssest automatiseerimisest üleminekul integreeritud süsteemile. Kuna viimase tuumaks on arvutikataloog, siis tavaliselt selle koostamisest alustataksegi. Järgnev ülevaade kataloogimise automatiseerimise algusest Tartu Ülikooli Raamatukogus on antud eelkõige kataloogija pilgu läbi.

Konkreetsed eeltöid üleminekuks kaartkataloogimiselt arvutikataloogimisele alustasime osakonnas 1993. aastal. Palju tähelepanu pöörasime standardiseerimisele, mis on automatiseerimise alus ja tagab tulevikus erinevate süsteemide ühildatavuse. /1/

1. Kirje koostamisel hakkasime kasutama uut rahvusvahelist standardit ISBD 1987.

2. Kui Eestis otsustati valida bibliikirjete vahetusformaadi aluseks FINMARC, tõlkisime selle eesti keelde.

3. Vastavalt ERÜ automatiseerimise töögrupi poolt 1994. aastal koostatud "Eesti vahetusformaadi" projektile (so. lihtsustatud variant FINMARCist) määratlesime arvutiandmebaasi raamatukirjes kajastuva andmestiku, mis oli abiks andmetöötlusosakonnale nende sisestusprogrammi koostamisel. (Nimetatud projekt on läbi arutatud ERÜ kirje ning liigitamise ja märksõnastamise töögrupis, kuid vajab veel eriti infokoodide kasutamise osas täpsustamist.

4. Meie liigitajad asusid koos teiste teadusraamatukogude liigitajatega tõlkima uusimaid ingliskeelseid UDK liigitustabeleid.

5. Osakonda seati üles kolm terminali ja üks personaalarvuti koos printeriga, et alustada arvutikoolitust. /2/

1994. aasta kevadel valmis andmetöötlusosakonnas raamatukirjete **sisestusprogramm**. Ligi pool aastat kestis selle katsetamine ja täiendamine, mille käigus pikendati andmevälju, lisati abitabelid jne. Andmetöötlusosakond koostas kataloogijatele ka lühikese tööjuhendi.

1994. aasta novembris alustasime saabuvate raamatute (ladina tähestikus, ilmunud 1994. aastal) **kataloogimist arvuti abil** (*original cataloguing*). /3/

Kõige rohkem muutis see **liigitajate** tööd. Neil tuleb lisaks kirjanduse liigitamisele see ka märksõnastada. Et üldist eestikeelset märksõnastikku uuema kirjanduse kohta veel ei ole, otsustasime luua nõ. töö käigus vabalt märksõnastades oma raamatukogu **tesauruse**. Selleks on eraldi abiprogramm. Kirjele saab lisada kuni kümme **märksõna**, mis peavad olema eelnevalt tesauruses registreeritud. Märksõnad on antud eesti ja inglise keeles, igal **raamatul** võib olla piiramatult arv tähenduselt laiemaid ja kitsamaid märksõnu. Veel saab rakendada viiteid "kasuta" (viitab sünonüümilt deskriptorile) ja "vaata ka" (viitab assotsiatsioonisuhtes olevale märksõnale).

Märksõnastiku haldamiseks tuleks seda programmi siiski mõnevõrra täiendada. Praegu ei saa märksõna juurde kirjutada selgitavat vm. teksti, mis ei kuulu märksõna koostisse (*scope note*). See oleks aga väga vajalik niisuguse mitme eri inimese poolt pidevalt täiendatava tesauruse loomisel. Teiseks oleks hea, kui

märksõnale saaks viitena lisada **raamatul** UDK indeksi. Samuti on veel lahendamata probleem, kuidas märksõnu automaatselt suurematesse süsteematailistesse klassidesse rühmitada.

Märksõnadeks valitakse võimalikult lihtsaid mõisteid väljendavad sõnad - deskriptorid. Deskriptor-tüüpi märksõnu peetakse automatiseeritud andmebaasides üldiselt paremaks kui *subject heading*-tüüpi märksõnu. Meie lähtusime valiku tegemisel järgmistest kaalutlustest:

1. Deskriptorite kasutamisel koguneb märksõnastikku vähem sõnu, nende moodustamise reegleid saab selgemini määratleda ja nende vaheliste seoste kujundamist paremini kontrollida.

2. Deskriptorid sobivad paremini märksõnade kombineerimiseks otsingu ajal; kõiki vajalikke kombinatsioone nagunii ette anda ei saa.

3. Koostatavas Eesti üldises märksõnastikus, mida tahame oma raamatukogus rakendada nii palju, kui võimalik, kasutatakse samuti deskriptor-tüüpi märksõnu.

Vormilt võib deskriptor olla liht- või liitnimisõna, omadus- ja nimisõna ühend või harva kahe nimisõna ühend. Õnneks ei teki meie tesauruses suure ja väikese algustähe probleemi, sest märksõnades kasutatakse ainult suurtähti.

Tesauruse mahtu on raske ette ennustada. On avaldatud arvamust, et universaalsed teadusraamatukogud vajavad umbes 100 000 märksõnalist tesaurust. Näiteks Amsterdamis ülikooli raamatukogu märksõnastikku olevat juba esimese aastaga kogunenud 80 000 terminit, kui võtta sünonüümid ja kvaasisünonüümid ka arvesse.

Otsing hakkab toimuma loodava tesauruse kaudu. Kirjed on seotud otse ilma vaheregistrita vastavate märksõnadega tesauruses. Kuidas arvuti lugejat aineotsingul täpselt abistama hakkab, ei oska veel praegu öelda, sest lugejale mõeldud kasutajaliides on välja töötamata. Igal juhul peab lugeja päringu formuleerima küllalt palju järelkoordinatsiooni teel.

Liigiindekseid võib lisada igale kirjele kuni kümme. Eesti liigitamise standardi alus on UDK masinloetav väljaanne 1993. aastast - *Master Reference File* (MRF), mille järgi koostatakse UDK eestikeelne väljaanne. /4/ MRF erineb suuresti UDK eelmistest väljaannetest (näit. pealiigis 8). Nüüdsest meie raamatukogus rakendatavad indeksid on kõik vastavuses MRFIGa.

UDK indeksid sisestatakse arvutikataloogi tervikuna, st. määrajaid ei eraldata põhiindeksist. Kirjetele antud indekseid on praegu võimalik vaadata ainult kirje juures, registrit nendest ei teki. Seepärast ei saa teha ka otsingut UDK järgi. Selles suhtes tuleb programmi kindlasti täiendada, sest liigitajal oleks aeg-ajalt vaja lehitseda mingi kindla indeksi saanud kirjeid. Nagu näitavad teiste maade kogemused, ei saa UDK-otsing lugejate seas populaarseks enne, kui pole loodud allandmebaase või vähemalt trükiväljaandeid liigitustabelitest ja/või ainelooendist, mis aitaksid lugejal leida õige indeksi. Need aineotsingu abivahendid tuleb teha arvutikataloogi kasutajale kättesaadavaks võimalikult ruttu.

Vähem muutusi tõi arvutikataloogimine kirje koostajate töösse. Kuna meie raamatukogu integraalse ülesehitusega süsteem INGRID lubab kord süsteemi sisestatud andmeid kasutada mitmete raamatukogutöö ülesannete täitmiseks, siis kasutatakse kirje koostamisel komplekteerijate poolt kirjanduse arvelevõtmiseks koostatud lühikirjet. Kataloogija täiendab kirjet, lisab ilnumismaa ja

-keele koodi, täiendkirjete pealdised koos rollikoodidega ja võtmesõnad (dokumendi pealkirjas või alapealkirjas olevad sisu avamiseks olulised nimisõnad dokumendi keeles). Kataloogija sisestab ka raamatu alalise ja ajutise kohaviida ning trükib seejärel spetsiaalsele rebitavale kaardipaberile kataloogikaardi. Kataloogijate arvates on kirje sisestusprogramm õnnestunud. Kuigi arvutikataloogimine nõuab meie kataloogijalt äärmist täpsust ja täit tähelepanu, on arvutikirjet lihtne koostada. (Näiteks 6. jaanuari 1995 seisuga oli 108 kirje sisestamisel tekkinud ainult kaks probleemi, mis osutusid kergesti lahendatavateks.) Kirje valmib suhteliselt kiiresti, sest kataloogija peab täitma ainult kaheksa ekraanilehekülge, kus on antud meile kui teadusraamatukogule kohustuslike kirjeelementide nimetused eesti keeles. Vajadusel programmi täiustatakse. Kirjed arhiveeritakse iga päev andmetöötlusosakonnas.

Kataloogimise automatiseerimise järgmise etapi planeerimisel peame otsustama

- 1) missuguste saabuvate dokumentide kirjeid hakata järgmisena arvutisse sisestama,

- 2) kuidas peaksid bibliograafilised andmebaasid olema seotud arvutikataloogiga,

- 3) missuguseid kaartkatalooge hakata millises järjekorras arvutisse sisestama,

- 4) kuidas oleks võimalik kujundada ülikooli raamatukogu arvutikataloogist ülikooli kõikide allüksuste raamatuvara koondkataloog.

Lugeja vajadusi silmas pidades on kataloogimisosakonnal esmajärjekorras plaanis hakata arvutikataloogi täiendama aastatel 1990-1994 ilmunud kirjandusega ning seejärel ringluses olnud ja lugejate poolt tagastatud kirjandusega (ilmunud 1945-1989), mis näitaks ühtlasi komplekteerimise puudusi. Tagastatud kirjanduse ümbereinventeerimine võimaldab ka teha fondis inventuuri.

Kuna kogu sisestatav kirjandus vajab märksõnastamist, kirjed ja UDK indeksid korrigeerimist, siis tuleb hästi läbi mõelda, missugune kaartkataloogi töötlemise tehnoloogia oleks kõige otstarbekam ja ökonoomsem, et luua **retrospektiivset andmebaasi** (*retrospective cataloguing*). /5/

Kataloogijate töö lihtsustamiseks ja kiirendamiseks kasutatakse maailmas juba ammu kataloogimiskeskustes koostatud kirjeid, mis kantakse teistest süsteemidest kohalikku andmebaasi (*copy cataloguing*). /6/ Eesti raamatukogude automatiseerimise põhimõtete kohaselt hakkab ülikooli raamatukogu täitma arvutikataloogimiskeskuse ülesandeid vastavalt oma profiilile.

Selleks, et meie kataloogimiskeskuste töö sujuks, peaksime oma tegevust praegusest rohkem kooskõlastama. Eelkõige tuleks meil

- lõpule viia UDK liigitustabelite tõlkimine eesti keelde,

- luua universaalne eestikeelne tesaurus,

- anda välja ühtsed eestikeelsed kirjereeglid,

- töötada välja täiendkirjete ja võtmesõnade valiku printsiibid. /7/

Kui me aga tahame tulevikus integreeruda ühtsesse ülemaailmsesse süsteemi, peaksime juba praegu hakkama mõtlema ka kollektiiv- ja üksikautori nimede kasutusviisi ühtlustamisele kirjes rahvusvahelise standardi alusel. /8/ Peavad ju meie kirjeid mõistma ka välislugejad.

Kindlasti ootab kataloogimiskeskusi ees rida keerulisi probleeme. Suures osas aitaks neid lahendada ühtse töökindla andmetöötlussüsteemi rakendamine, mille kasuks otsustaksid kõik raamatukogud.

Lugejauuringute põhjal võib väita, et ülikooli raamatukogu ajalooliselt väljakujunenud kataloogisüsteem on üsna keeruline ja infootsinguks kulub

palju aega, kuigi seda on pidevalt püütud optimeerida ja teatmestut täiustada. Ilmselt aasta või paari pärast saame võimaldada lugejatele juurdepääsu arvutikataloogile ja täitub nende unistus leida kogu vajalik kirjandus võimalikult kergesti ühest kataloogist. Meie koostatud korrektne täiskirje peaks tagama, et lugeja leiab teda huvitava info võimalikult mitme tunnuse järgi kiiresti ja ammendavalt.

Niisiis oleme asunud kataloogimise automatiseerimise pikale teele, mis peaks meid 21. sajandil viima täielikult automatiseeritud kataloogimiseni (so. inimese vahetu osavõtuta, ainult tema järelvalve all). /9/ Kuigi see tundub utoopilisena, võib erialakirjanduse põhjal öelda, et kataloogimise täieliku automatiseerimisega tegeldakse. Näiteks otsitakse võimalusi, kuidas saaks arvuti kataloogida dokumenti kodeeritud tiitelile järgi ja märksõnastada sisukorra või kokkuvõtte järgi. Räägitakse isegi kataloogimise kriisist ja sellest, et kataloogimine kui selline on surev kunst.

Sajandi algul läksime sedelkataloogidelt üle kaartkataloogidele ja nüüd, sajandi lõpus oleme minemas kaartkataloogidelt üle arvutikataloogile. Kataloogi vorm ja kvaliteet muutub, kuid kataloogijaks jääb ikka vaid inimene - subjektiivne otsustaja. Niisiis jääb arvutikataloogi kommunikatiivsus ikkagi sõltuma kataloogija bibliograafilistest teadmistest, ainevaldkonna tundmisest ja keeleoskusest.

Artikli põhiseisukohad on ette kantud Tartus 5.-6. dets. 1994 peetud Tartu Ülikooli Raamatukogu teaduskonverentsil "Raamatukogu osa teaduslikus kommunikatsioonis"

KIRJANDUS

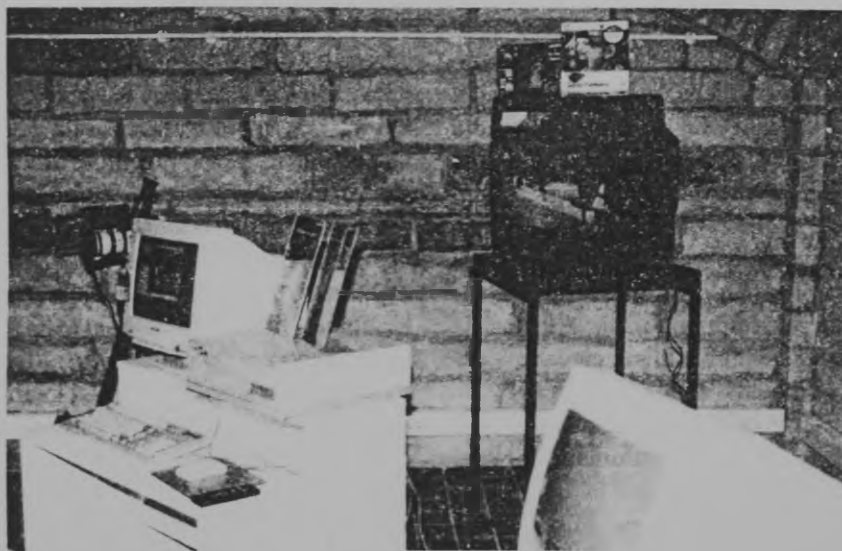
1. IFLA kataloogimissektsioon ühtlustamas eri maade bibliokirjeid / Refereerinud J. Järs // *Uut Välismaa Raamatukogudes*. 1992. - 7. Lk. 3-7.
2. Astrodata arvutiaabits / E. Saar, H. Ibrus, K. Annuk jt. 2., parandatud ja täiendatud tr. Tõravere, 1994. - 127 lk.: ill.
3. Carpenter, M. The narrow, rugged, uninteresting path finally becomes interesting : A review of work in descriptive cataloguing in 1991 with trail marks for further research // *Library Resources and Technical Services*. - 1992. - Vol. 36, N 3. P 291-315.
4. Kirsal, A.-M. Liigitamise ja märksõnastamise ühtsus raamatukogudes on võimalik // *Raamatukogu*. - 1994. 2. - Lk. 12-13.
5. Kaartkataloogide skaneerimistehnoloogial põhinev retrospektiivne konversioon / Refereerinud J. Järs // *Uut Välismaa Raamatukogudes*. - 1992. 7. Lk. 12-20.

6. LeBlanc, J. D. Cataloguing in the 1990s : managing the crisis (mentality) // *Library Resources and Technical Services*. 1993. - Vol. 37, N 4. P. 423-433.

7. Gattermann, G. Kann den Online-Benutzerkatalog unser Katalogproblem lösen? : Erfahrungen aus dem OPAC-Projekt der Universitätsbibliothek Düsseldorf // *Zeitschrift für Bibliothekswesen und Bibliographie*. 1990. S.-H. 50. S. 81-92.

8. Management and use of name authority files: personal names, corporate bodies and uniform titles / M. Beaudiquez, F. Bourdon; IFLA and Institutions et al. München etc., 1991. 90 p. (UBCIM publications; N. S., Vol. 5).

9. Davies, R. Expert systems and cataloguing // *The application of expert systems in libraries and information centres*. London etc., 1992. P 133-166.



Eesti Raamatukoguhoidjate Ühing

EESTI RAAMATUKOGUHOIDJATE ÜHINGU

AASTARAAMAT

7

1995

Tallinna Tehnikaülikooli kirjastus

Tallinn 1996

SISUKORD * CONTENTS

Saateks	5
Preface	6

TEADUSLIKUD KESKRAAMATUKOGUD

<i>Konrad Kikas</i>	Teaduslikud keskraamatukogud mujal ja meil	7
	Eesti teaduslikud keskraamatukogud (TKR) täna	23
	1995. aasta statistikat	36
<i>KHM määrus</i> <i>12.12.1994</i>	Teadusliku keskraamatukogu tegevuse põhimõtted	39
<i>KHM määrus</i> <i>19.01.1995</i>	Teaduslike keskraamatukogude ja nende ainevaldkondade loetelu	40

AKTUAALSEL TEEMAL

<i>Jüri Järs</i>	Tiigrihüpe ja raamatukogude automatiseerimine A. D. MCMXCV	42
<i>Sirje Nilbe</i>	Tesauruse mitu palet	51
<i>Arda-Maria Kirsell</i>	UDK eestikeelse väljaande saamisloost	68
<i>Tiiu Pihlakas</i>	Maaraamatukogud täna	76

ARVUSTUSI. ARVAMUSI

<i>Marje Aasmets</i>	Raamatukogunduslikke jadaväljaandeid aastail 1990-1995	80
<i>Villu Tamul</i>	Arvo Tering - mitmekordne laurea	90

AJALOOLIST

<i>Ivi Tingre</i>	Eesti raamatukogud ja nõukogude raamatukogu- poliitika	93
-------------------	---	----

TÄHTPÄEVI

<i>Heino Räim</i>	Helene Kulpa vaatame koos tagasi	101
<i>Ingrid Loosme,</i> <i>Linda Kivistik,</i> <i>Maare Kümnik,</i> <i>Aime Naanuri</i>	Mõttekilde Kaja Noodlast kui õpetajast	105

Märksõnastus muutumas

Omalooodud elektrooniliste andmebaaside, kaasa arvatud arvutikataloogide lisandudes on meie raamatukogudes tarvis ümber mõtestada senised teavikute sisu avamise meetodid. Süstematiseerimine kaotab arvutis palju neist eelistest, mis tal on kaartkataloogis, sest süsteemi pole ekraanil näha. Indeksi järgi otsimiseks peab hästi tundma UDK notatsiooni, mida aga ei saa tavalugejalt loota. Lugejale on tähtis, et ta saaks oma päringu infokeelde tõlkimisega ise hakkama. Seepärast eelistab ta kindlasti märksõnu.

Märksõnastus kui trükiste sisu avamise meetod on meil üsna hästi tuntud. Kuigi suured universaalsed sisukataloogid on üles ehitatud süstemaatiliselt UDK järgi, leidub igas raamatukogus kindlasti mõni märksõnakataloog või -kartoteek. Märksõnastikud nende kataloogide ja kartoteekide jaoks on kujunenud töö käigus, vastavalt teavikute iseloomule ja hulgale. Nad ei ole mõeldud rakendamiseks mõne teise kogu puhul. Eesti keeles on olemas ka metoodiline abivahend märksõnastamise ja märksõnakataloogi koostamise kohta Helmi Masingu "Märksõnaliigitus" [12].

Kuivõrd tõhus on traditsiooniline märksõnastus arvutikataloogis? Lugesdes erialakirjandusest, kui raske on lugejatele aineotsing Kongressi Raamatukogu märksõnastiku abil indekseeritud kataloogidest, tuleb tõdeda, et arvutikataloog vajab teist tüüpi märksõnastikku ja teistsugust märksõnastamise metoodikat. Inversiooni, liitsõnade poolitamise, allmärksõnade, viidete jms abil hoolikalt kujundatud märksõnarubriikide ja -komplekside puhul ilmneb seesama puudus, mis UDK juures - rubriikide liigendust ja järgnevust pole arvutis näha. Võib öelda ka nii: kui kaartkataloogis on keskseks ühikuks rubriik, siis arvutikataloogis on selleks konkreetse teaviku kohta koostatud kirje. Seda ühte kirjet saab otsida mitmesuguste tunnuste järgi: autor, pealkiri, ilmumisaasta, ilmumiskoht, märksõna, liigiindeks. Põhimõtteliselt on kirje leitav iga selles esineva sümbolijärjendi (sõne) kaudu, seega pole aineotsinguks vaja tingimata märksõna või indeksit kasutada, vaid sobiva teose võib leida ka näiteks pealkirjas esineva sõna või sõnaosa järgi. Nn võtmesõna- ehk vabasõnaotsing on eriti levinud artiklite referaatandmebaasides, mis sisaldavad kirjutiste kokkuvõtteid, ja teksti-

andmebaasides, kus elektroonilisel kujul on olemas kogu teaviku tekst. Enamik raamatukogukatalooge sisaldab siiski vaid traditsioonilisi kirjelemente, mis võtmesõnaotsinguks eriti palju ei paku, nii et sisu avamiseks tuleb kirjetele lisada täiendavat informatsiooni.

Kõige objektiivsem ja odavam oleks märksõnastada teaviku tekstis endas esinevate terminite abil. See on vaba märksõnastamise üks variant. Aga see teeks info taasleidmise raskeks: ammendava tulemi saamiseks peaks läbi proovima mitu sünonüümset väljendit mitmes keeles. Mittetekstiliste teavikute, nagu piltide ja helisalvestiste sisu tuleb igal juhul verbaalsesse keelde tõlkida.

Lühidalt infotehnoloogia praegusel arenguetapil oma katalooge arvutisse viies ei saa me vältida ohjatud märksõnastamist (*controlled indexing*) raamatukogus hoitavate teavikute mitmekesisuse ja kataloogikirje iseloomu tõttu.

Milline võiks olla infokeel, mis sobiks kõige paremini arvutikataloogide märksõnastamiseks? Praktika on näidanud, et kõige paindlikumaid otsivõimalusi pakub meetod, kui teaviku sisu väljendada rühma iseseisvate lihtmõisteid tähistavate loomuliku keele terminite kaudu (koordinatiivindekseerimine). Siis pole märksõnastamine enam teatud hulgale kirjetele sobiva "ühtluspealkirja" leidmine (vrd ingl k *subject heading!*), vaid iga teavik võib saada kordumatu, tema sisu kõige täpsemalt avava märksõnaloendi. Antud infokeeles lubatud sõnavara on võimalik esitada erineval viisil - lihtsast märksõnaloendist põhjalikult läbitöötatud **tesauruseni**. Kõige populaarsemad infokeeled ongi viimastel aastakümnetel olnud mitmesugused tesaurused.

Meilgi räägitakse üha enam vajadusest üldise eesti tesauruse järele, et ühtlustada sisupeegeldust eri raamatukogudes. Üldist tesaurust koostatakse Rahvusraamatukogus. Tartu Ülikooli raamatukogus aga kujuneb alates 1994. aasta novembrist oma arvutikataloogiga integreeritud tesaurus. Eesti keeles on olemas ka üks erialatesaurus **URBISe** märksõnastik fennougristikakirjanduse märksõnastamiseks.

Tesauruse mõistet on informatsiooninduses kasutatud ligi nelikümmend aastat. Kõige laiemalt võib tesaurust defineerida kui sõnavaraohje (*vocabulary control*) vahendit automatiseeritud infosüsteemides. Määratlus on nii lai sellepärast, et sõnaga *tesaurus* on erialakirjanduses üsna pillavalt ringi käidud: selleks on nimetatud nii kõige lihtsamat deskriptorite loendit,

mida mõnes andmebaasis kasutatakse, kui ka näiteks intelligentse süsteemi leksikaal-semantilist teadmusbaasi. Kõige sagedamini räägitakse siiski tesaaurusest kui märksõnastikust infotöötluste ja -otsingu jaoks, mis sisaldab loomulikult keelest valitud sõnu ja sõnaühendeid ning nende vahelisi semantilisi seoseid.

Tesaauruse arenguloost

Tesaauruse idee võtsid infoteadlased üle leksikograafiast* Leksikograafias nimetatakse tesaauruseks sellist tüüpi **sõnaraamatut**, kus sõnad on korrastatud sisuliselt, mitte alfabeetiliselt. Ühte on koondatud kindla mõiste või ideega seotud sõnad ja väljendid. Mõistelisi sõnaraamatuid on koostatud juba meie ajaarvamise esimestel sajanditel. Näiteks pärineb 2. sajandist Polluxi vanakreeka sõnaraamat "Onomastikon" ja 6. sajandist sanskriti keelevaramu, Amarasimha "Amara-kosha"

Kõige tuntum sõnaraamat-tesaurus on arvatavasti P. M. Roget "Thesaurus of English Words and Phrases", mille esimene väljaanne ilmus 1852. aastal. Hiljem on seda uuendatud ja taasavaldatud eri mahus, kuid sõnaraamatu struktuuri pole muudetud. Roget tesaaurus koosneb kahest osast: mõistepiirkonniti esitatud põhiosa ja tähestikregister. Leksikoniüksuseks on "idee" sõna, mis märgib teatud kompleksset tähendust või semantilist välja. "Ideed" on süstematiseeritud kaheastmelise klassifikatsiooniskeemi järgi ja nummerdatud. Iga "idee" juures on loetletud sõnu ja väljendeid, mis tähenduslikult selle "ideega" seostuvad. Registri abil leiame näiteks sõna *library* mõisterühmast 589. *Book*, milles peale selle sõna on veel poolteist tihedat veergu raamatu, kirjastamise ja kirjutamisega seotud sõnu [13, 245].

Eesti keele sõnastikest võiks tesaauruseks nimetada Andrus Saareste "Eesti keele mõistelist sõnaraamatut" Meie keeleteadlaste seas on idanenud mõte luua elektrooniline eesti keele tesaaurus, mis sisaldaks rikkalikku infot sõnade tähenduse ja omavaheliste seoste kohta [17].

Keeletesauruste ülesanne on rikastada ja täpsustada inimeste keelekasutust, aidata neil luua hea stiiliga tekste. Sõnaraamat juhib meid "ideesõnadelt" mõistetelt, millest tahame rääkida või kirjutada teksti-

* Sõna ise pärineb kreeka keelest (*thēsauros*) ja tähendab algselt aaret või varakambrit.

sõnadeni, mis väljendavad meie mõtet kõige paremini. Teavikute töötlemisel infosüsteemides on vastupidine eesmärk: sisult sarnastele, kuid erinevaid sõnu sisaldavatele tekstidele tuleb leida mingi ühine "ideesõna". Kuigi need kaks tegevust on vastupidise suunaga ning keeletesaurused ei saa asendada infotesaurusi ja vastupidi, on nad piisavalt sarnased selleks, et abivahendid nende protseduuride jaoks võiksid kanda ühist perekonnanime.

D. A. Krooksi ja F. W. Lancasteri väitel hakkasid spetsialistid tesauruse rakendamise üle infootsingus arutlema 1956. aastal Inglismaal. Esimesed artiklid sel teemal ilmusid 1957. aastal Ameerikas [7, 328].

B. C. Vickery on kirjeldanud nelja erinevat ettekujutust **infotesaurusest**, mis 1950. aastate lõpul esitati [16]. Neist H. P. Luhni oma lähtus automatiseeritud indekseerimise probleemist. Ta pani ette grupeerida tähenduselt seotud sõnad mõisterühmadesse (*notional families*) ja tähistada iga rühm koodiga, näiteks numbriga, nii nagu on Roget sõnaraamatus. Süsteem tuvastab tema mälus olevate sõnade esinemise tekstis ja lisab kirjele vastava(d) koodi(d).

C. L. Bernier ja K. F. Heumann nägid tesauruses vahendit kasutaja ja otsisüsteemi sõnavara ühteviimiseks. Tesaurus pidi nende arvates sisaldama piiratud arvul definitsioonidega varustatud märksõnu, mida süsteemis on kasutatud, ja suure hulga "tekstisõnu", mida võiks kasutada infonõudleja. Tekstisõnadelt on viited märksõnadele ja vastupidi.

Kolmas lähenemisviis, mille esitas E. Wall, pööras tähelepanu celkõige märksõnade vaheliste semantiliste suhete väljendamisele. Wall eristas sünonüümia-, üldise-üksiku ja assotsiatsioonisuhet ning tesaurus tema nägemuses oli märksõnade tähestikloend, kus iga märksõna all olid viited sellega mingis seoses olevatele märksõnadele.

J. H. Healdi ettekujutus tesaurusest sarnanes Walli omaga. Tesaurus pidi olema viidetega varustatud märksõnade loend, kuid antud märksõna suhtes üldisema mõiste ("kategooria") paigutas Heald märksõna järele sulgudesse.

Tänapäeva standardne tesaurus sarnaneb kõige rohkem Walli variandiga, mille ülesanne oli aidata nii märksõnastajal kui ka infonõudlejal leida lisasõnu kas siis teaviku indekseerimiseks või oma päringu esitamiseks. Kuid kasutatud ja edasi arendatud on ka Luhni ning Bernier ja Heumanni ideid.

E. Walli põhimõtetele toetudes valmis esimene praktilises töös

rakendatud teaurus 1959. aastal Ameerika infotalituses *Engineering Information Center of E. I. Dupont de Nemours and Company*. Esimesed publitseeritud teaurused olid "Thesaurus of ASTIA Descriptors" (1960) ja "Chemical Engineering Thesaurus" (1961), mis ilmusid samuti Ameerikas [7, 326-328]. Sellest ajast peale on teaurusi koostatud paljudes keeltes paljudes maades, üks- ja mitmekeelseid, kitsalt erialaseid ja laiema ainehaardega, alfabeetilise ja süstemaatilise ülesehitusega. Nende kohta saab andmeid teatmeteostest [3; 4; 15] ja ECHO andmebaasist THESAURI. ISO on kehtestanud nõude, et igast publitseeritud teaurusest tuleb üks eksemplar saata rahvusvahelisse arvelduskotta (*clearinghouse*) Torontosse (inglisekeelsed ja inglise keelt sisaldavad mitmekeelsed teaurused) või Varssavisse (teaurused teistes keeltes).

Tesauruse ülesehitus

Tesauruse koostamist reeglistavad rahvusvahelised standardid ISO 2788 [5] ja ISO 5964 [6]. Enamik rahvuslikke standardeid lähtub nendest, kuid arvestab ka oma maa traditsioone ja keele eripära. Soomekeelse teauruse standard valmis 1988. aastal [14]. See võiks olla üks eeskuju Eesti standardile, kui seda kunagi tegema peaks hakatama. Üldtunnustatud käsiraamat, mis toetub kehtivatele standarditele, kuid annab teaurusega seotud probleemidest palju avarama pildi, on J. Aitchisoni ja A. Gilchristi "Thesaurus Construction: a Practical Manual" [1].

Tesauruse **struktuuri** moodustavad terminid ja nende vahelised seosed. Terminid väljendavad märksõnastiku ainevaldkonna olulisi mõisteid. Loomulikule keelele, ka oskuskeelele on omane sünonüümia: üht mõistet saab väljendada mitut moodi (*kaalanalüüs ~ gravimeetriline analüüs*). Tesauruse ülesanne on piirata infosüsteemis kasutatavate terminite hulka ja üks meetod selleks ongi sünonüümsete väljendite kokkuviiimine (vt ekvivalentsusseos allpool). Seega sisaldab teaurus kahesuguseid termineid: märksõnu ja viitetermineid, mis järjestatakse tähestikuliselt ühtsesse nimestikku.

Tesauruses kajastub ka teatud hulk mõistete vahelisi semantilisi suhteid. Neid väljendatakse viidetega termini juurest teistele terminitele. Tavaliselt eristatakse kolme semantilist seost: ekvivalentsusseos, hierarhiaseos ja assotsiatsiooniseos. Neid võib välja tuua ka rohkem, kuid

tuleb silmas pidada, et tesauruse struktuur ei muutuks ülearu keeruliseks.

Ekvivalentsusseos ehk samaväärsusseos moodustatakse selliste terminite vahele, mis tähistavad üht ja sama mõistet või niivõrd lähedasi mõisteid, et märksõnastamisel ja infootsingul ei ole otstarbekas neid lahus hoida. Üks ekvivalentide klassi liige valitakse märksõnaks (*preferred term, descriptor*), teine läheb tesaurusesse viiteterminina (*non-preferred term, entry term, non-descriptor*). Ühel märksõnal võib olla mitu viiteterminit. Viiteterminid on soovitatav trükkida teistsuguses šriftis kui märksõnad.

arstiteadus

kasuta MEDITSIIN

MEDITSIIN

asendab *arstiteadus*

Ekvivalentsusseoses terminid võivad olla täielikud või osalised sünonüümid (kvaasisünonüümid). Täielikud sünonüümid on näiteks omasõna ja võõrsõna

finantspoliitika

kasuta RAHANDUSPOLIITIKA

RAHANDUSPOLIITIKA

asendab *finantspoliitika*

• ühe mõiste erinevad väljendused eri teadusaladel

kõnehäired

kasuta KÕNEPUUDED

KÕNEPUUDED

asendab *kõnehäired*

• vananenud ja uus termin

telefoniraamatud

kasuta TELEFONIKATALOOGID

TELEFONIKATALOOGID

asendab *telefoniraamatud*

lühend ja täisnimetus

USA

kasuta AMEERIKA ÜHENDRIIGID

AMEERIKA ÜHENDRIIGID

asendab *USA*

Eestikeelse terminoloogia praeguses seisus tuleks kõne alla ka levinud barbarismi ja kinnistumata omatermini sidumine:

public relations

kasuta ÜLDSUSSUHTED

(AVALIKKUSSUHTED?)

(SUHTEKORRALDUS?)

ÜLDSUSSUHTED

asendab *public relations*

Kvaasisünonüümid on sellised terminid, mille tähendus kattub osaliselt ja mida on otstarbekas käsitada sünonüümidena. Näiteks saab niimoodi piirata tesauruse ainevaldkonna äärealadele jäävate või raskesti defineeritavate märksõnade hulka.

seiklusromaanid

kasuta PÕNEVUSROMAANID

PÕNEVUSROMAANID

asendab *seiklusromaanid*

Kvaasisünonüümidena saab ühendada ka antonüüme, kui need sobiksid märksõnadena tavaliselt ühele ja samale teavikule:

kirjaoskamatus

kasuta KIRJAOSKUS

KIRJAOSKUS

asendab *kirjaoskamatus*

Kui mingit ainelõiku ei ole mingil põhjusel, nt materjali vähesuse või kõrvalise tähtsuse tõttu mõttekas sügavuti märksõnastada, aga oleks

vaja näidata, milliseid spetsiifilisemaid mõisteid see hõlmab, kasutatakse samuti ekvivalentsusseost.

MÄESUUSATAMINE

asendab *kiirlaskumine*
slaalom
sööstlaskumine

kiirlaskumine

kasuta MÄESUUSATAMINE

slaalom

kasuta MÄESUUSATAMINE

sööstlaskumine

kasuta MÄESUUSATAMINE

Sisuliselt on siin tegemist hierarhiasuhtega, kuid et märksõnaks on valitud ainult laiem termin, tuleb see vormistada ekvivalentsusseosena.

Hierarhiasuhe on selliste mõistete vahel, millest üks sisaldub tervikuna teises, moodustades sellest osa. Rangelt võttes ja standardi järgi on hierarhiaseos lubatud ainult selliste terminite vahel, mis kuuluvad ühte semantilisse kategooriasse, st mõlemad tähistavad kas nähtust, omadust, tegevust, teadusdistsipliini vms. Hierarhias

MUUSEUMID

kitsam KODULOOMUUSEUMID
MUUSEUMITÖÖTAJAD
MUSEOLOOGIA
TEATRIMUUSEUMID

ei kuulu kõik terminid ühte kategooriasse ning MUUSEUMITÖÖTAJAD ja MUSEOLOOGIA ei saa olla termini MUUSEUMID suhtes "kitsamad". Omaette küsimus on, kas sellised ranged põhimõtted hierarhiaseoste moodustamisel end otsingu seisukohalt alati õigustavad.

Hierarhiasuhte alaliigid on geneeriline ehk soo-liigisuhe, partitiivne ehk osa-tervikusuhe ja üldmõiste-üksikjuhtumi suhe.

Geneerilise seose puhul on laiemaks soomõiste ja kitsamaks liigimõiste. Tavaliselt on ühel soomõistel mitu alluvat liigimõistet.

LÄÄNEMERESOOME KEELED

laiem SOOME-UGRI KEELED

kitsam EESTI KEEL

ISURI KEEL

KARJALA KEEL

LIIVI KEEL

SOOME KEEL

VADJA KEEL

VEPSA KEEL

ELEKTRISEADISED

kitsam ELEKTRIMASINAD

ELEKTRIMÕÕTERIISTAD

ELEKTRONSEADISED

Partitiivse seose puhul väljendab laiem mõiste tervikut ja kitsam selle osa. Selliseid sõnarühmi, kus osa ja terviku suhe on kontekstist hoolimata püsiv, pole kuigi palju. Näiteks võib tuua

anatomilised struktuurid

SEEDEELUNDID

kitsam MAGU

MAKS

NEEL

SOOLED

*kohanimed

EESTI

kitsam HIIUMAA

IDA-VIRUMAA

JÕGEVAMAA

*teadused ja teadusharud

KEEMIA

kitsam ANALÜÜTILINE KEEMIA

BIOKEEMIA

FÜÜSIKALINE KEEMIA

ORGAANILINE KEEMIA

Üldmõiste ja üksikjuhtumi seosega (*instance relationship*) on tegemist siis, kui hierarhia moodustavad üldnimetus ja sellest kitsamate terminitena nimed.

MERED

kitsam BALTI MERI
KASPIA MERI
VAHEMERI

Kui nimed on otsustatud tesauruse põhiosast välja jätta ja lisada näiteks omaette loendina, siis sellist seost muidugi moodustada ei saa.

Assotsiatsiooniseost on raskem defineerida kui eelmisi seoseliike. Üldiselt esineb assotsiatsiooniseos selliste terminite vahel, mis on mõisteliselt küll tihedalt seotud, aga mitte hierarhiasuhtes, samuti pole nad ühe ja sama ekvivalentide klassi liikmed.

SEEDEELUNDID

vaata ka SEEDIMINE

SEEDIMINE

vaata ka SEEDEELUNDID

MAJANDUSLIIT

vaata ka AVATUD MAJANDUS
MAJANDUSINTEGRATSIOON

Assotsiatsiooniseoses termineid esineb rohkesti humanitaar- ja sotsiaaalaladel, kus kategooriate ja hierarhiate moodustamisest on tihti parem loobuda, et vältida subjektiivseid otsuseid.

Märksõnaartikkel näeb tesauruses välja nii (alati ei pea kõik väljad olema täidetud ja igal väljal võib olla mitu sõna):

RAHANDUSPOLIITIKA

asendab *finantspoliitika*
laiem MAJANDUSPOLIITIKA
kitsam MAKSUPOLIITIKA
vaata ka RIIKLIK RAHANDUS

Viitetermini puhul ei anta tavaliselt muud informatsiooni kui viide märksõnale:

finantspoliitika

kasuta RAHANDUSPOLIITIKA

Käesolevas kirjutises on seoseid tähistatud sõnadega (*kasuta, asendab* jt), sest meil pole veel kindlaks kujunenud tava, milliseid lühendeid või sümboleid kasutada. Rahvusvaheliselt on hästi tuntud ingliskeelsed lühendid, kuid reeglina võetakse lühendid tesauruse (dominant)keelest. Tesauruse standardites soovitatakse seoseid märkida keelest sõltumatute sümbolitega, kuid tundub, et neid kasutatakse siiski harva tähtlühendid jäävad paremini meelde. Järgnevas tabelis on kõrvuti ettepanekud eestikeelse tesauruse jaoks, ingliskeelsed lühendid ja ISO soovitatud sümbolid.

KAS	(kasuta terminit)	USE		→
AT	(asendab terminit)	UF	(used for)	=
LT	(laiem termin)	BT	(broader term)	<
KT	(kitsam termin)	NT	(narrower term)	>
VK	(vaata ka)	RT	(related term)	—

Eelkirjeldatud viisil esitatakse märksõnaartiklid sellises tesauruses, mille põhiosa on tähestikuline. Enamik tesaurusi ongi nii üles ehitatud. Näiteks võib tuua eesti märksõnastiku aluseks olevad väljaanded "Yleinen suomalainen asiasanasto" ja "SPINES Thesaurus". Süstemaatiline osa piirub neis suurte temaatiliste rühmade kaupa loetletud deskriptoritega. Juhtub kui süstemaatika on hästi läbi töötatud ja moodustab tesauruse tähtsama osa, jäetakse tähestikregistrist tavaliselt hierarhiaseosed välja. Selline on "Medical Subject Headings".

Tesaurus ja arvuti

Tesauruse koostamine ja edasine haldamine sisaldab rohkesti mehaanilist, kuid samas suurt täpsust ja inimeselt palju aega nõudvat tööd. Seepärast on tänapäeval üldine, et tesaurusi tehakse ja toimetatakse arvuti abil. Kasutatav tarkvara põhineb tavaliselt mõnel universaalsel baasiohjesüsteemil, mille tesaurust haldav asutus on kohandanud vastavalt oma vajadustele ja võimalustele. Kaubalisi programmpakette spetsiaalselt tesauruse jaoks on J. L. Milstedi hinnangul veel vähe saada, kuigi turg laieneb pidevalt [11, 390].

Kõige raskem on tesauruse juures seoste moodustamine ja kontroll. Muidugi ei suuda arvuti hinnata, kas seos on semantiliselt korrektne. See jääb inimese ülesandeks. Kuid sobiva tarkvara abil saab vältida vigu ja

võita aega seoste vormistamisel. Näiteks sõnade MOOTORSÕIDUKID, AUTOD, BUSSID, MOOTORRATTAD, SÕIDUKID sidumiseks hierarhiaseosega tuleks paberil teha täiendusi viie sõna juurest:

MOOTORSÕIDUKID
 laiem SÕIDUKID
 kitsam AUTOD
 BUSSID
 MOOTORRATTAD

AUTOD
 laiem MOOTORSÕIDUKID

BUSSID
 laiem MOOTORSÕIDUKID

MOOTORRATTAD
 laiem MOOTORSÕIDUKID

SÕIDUKID
 kitsam MOOTORSÕIDUKID

Tesauruse tarkvara kasutades piisab, kui sisestada vaid seoste üks pool (sõna MOOTORSÕIDUKID juurest), teine pool tekib automaatselt ja midagi ei unustata ära.

Kahepoolsete seoste genereerimine ühepoolse sisendi põhjal on miinimumnõue, millele peab vastama tesauruse koostamise tarkvara [11, 390]. Aga sellest ei piisa paberitööst loobumiseks. Kui juba arvuti abiks võtta, peaks programm võimaldama säilitada kogu tesauruse teksti, k.a märkused ja definitsioonid, sortida ja järjestada sõnu mitmesuguste tunnuste alusel, teha väljatrükke kogu tesaurusest või selle osast. Kui eesmärgiks on tesaurus publitseerida, peab olema võimalik kujundada andmebaasist hea tüpograafilise kvaliteediga fotoladu.

Masinloetav tesaurus on suurepärane abivahend **märksõnastamisel**, eriti kui tesaurus ja bibliograafiline andmebaas on omavahel ühendatud, s.o integreeritud. Hea programm ei lase märksõnastada terminitega, mis pole tesauruses autoriseeritud (nii antakse märku ka kirjavigadest). Samm edasi on võimalus valida kirjele märksõna ekraanile telli-

tud nimestikust ühe klahvilöögiga. Kui märksõnastaja juhtub valima viite-termini, asendub see automaatselt vastava deskriptoriga. Kasulik on näha ka terminite vahelisi seoseid, et leida kõige õigem(ad) märksõna(d). Sellise märksõnastussüsteemi eelis on vigade vältimine, kuid programmeerimisel tuleb hoolitseda, et terminite leidmine tesaarusest ja automaatne kirjele lisamine ei võtaks kokkuvõttes rohkem aega kui nende väljakirjutamine.

Andmebaasiga integreeritud tesaarus võimaldab parandada märksõna nii, et muutus peegeldub kohe kõigi nende kirjete juures, mis on selle märksõnaga indekseeritud. See on eriti tähtis juhul, kui tesaaruse loomine ja märksõnastamine toimuvad paralleelselt. Siis tuleb sageli parandusi teha. Ka hiljem on vaja asendada vananenud termineid uutega. Kui mingi sõna tesaarusest kustutada, peab see kustuma ka kirjete juurest. Aeganõudvam, kuid turvalisem on selline variant, kus kustutada saab ainult kasutamata märksõnu, s.t vastavad kirjed tuleb enne üle vaadata ja parandada, eemaldades soovimatu märksõna. Alles siis saab märksõna tesaarusest kustutada või muuta viiteterminiks.

Üsna palju on tegeldud probleemiga, kuidas masinloetava tesaaruse abil tõhustada **infootsingut**. Need võimalused, mida pakutakse märksõnastajale termini automaatne valimine nimestikust ja seoste vaatamine on kasulikud ka infonõudlejale, kuid otsitesaaruse saab kujundada ka sõltumata märksõnastamisest.

On hulk sidus- ja kompaktandmebaase, mille kasutaja saab lülitada tesaarusele, et otsida sobivaid märksõnu. Sageli on iga märksõna juures näha sellega indekseeritud kirjete arv, mille järgi kasutaja otsustab, kas on otstarbekas päringut kitsendada või laiendada. Parimad süsteemid pakuvad automaatset laiendust (*explode function*), mis tähendab, et soovi korral laieneb otsing kõigile antud märksõna suhtes kitsamatele mõistetele. Näib, et kõige mitmekülgsemad võimalused tesaaruse abil otsimiseks on info-pankade STN ja KRI/Dialog andmebaasides [8]. Nende hulka kuuluvad MEDLINE, GEOREF, INSPEC, ERIC jt tuntud andmebaasid, mida märksõnastatakse sama tuntud tesaaruste abil.

Raamatukogude siduskataloogides on otsitesaarust palju harvem rakendatud. Selle põhjused on eelkõige universaalsete tesaaruste vähesus ja asjaolu, et kataloogi kasutajaliides peab olema suunatud lõppkasutajale - sageli väheste kogemustega või lausa juhukülastajale, kellele tesaaruses

seilamine ja selle ehitusest arusaamine on liiga raske. Siinkohal pakub huvi M. Batesi käsitus otsitesaurusest (ta nimetab seda *end-user thesaurus*) [2].

Bates arvab, et enamik tesaurusi on koostatud ainult märksõnastajate vajadusi silmas pidades ega kõlba sellisena otsingu abivahendiks raamatukogukataloogis. Ta esitab järgmised otsitesauruse põhijooned:

Otsitesaurus sisaldab kõik sisu avamiseks kasutatud terminid, sh personaalia ja vabad märksõnad.

Andmebaasis kasutatud ja veel kasutamata, kuid võimalikud märksõnad on eristatavad.

Märkused (*scope notes*) on suunatud eelkõige lugejale, mitte indekseerijale. Nad sisaldavad rohkem definitsioone ja juhiseid terminite kombineerimiseks kui tavalises tesauruses.

Terminite vahelised suhted on väljendatud loomulikus keeles, mitte lühenditena.

Otsitesauruse viiteterminite (*entry vocabulary*) arv on väga suur. See on Batesi arvates otsitesauruse tähtsaim omadus. Viiteterminiteks võivad muuhulgas olla üldkeele ja kõnekeele sõnad ning sellised grammatilised vormid, mis ei vasta märksõnastuse reeglitele. Viiteterminid on kasutaseoses märksõnadega.

Otsitesaurus võib sisaldada termineid teistest, nii ohjatud kui vabadest infokeeltest. Eriti kasulik on siduda märksõnadega võtmesõnu.

Batesi otsitesauruse idee - lugeja ei tohi sattuda ummikusse ega jääda edasise tegutsemise juhiseita - lähtus vajadusest tõhustada aineotsingut LCSH abil märksõnastatud raamatukogukataloogides. Tesaurus pidi moodustuma poolautomaatselt andmebaasis leiduvatest elementidest. Kuid mitut siin esitatud põhimõtet tasuks arvestada juba tesauruse koostamise ja elektronkataloogimise algusjärgus.

1950. aastate lõpul esitatud infotesauruse "algkujudest" on otsitesaurus lähim Bernier' ja Heumanni ettekujutusele tesaurusest kui vahendajast süsteemi ja kasutaja vahel.

Et tesaurus on näidanud ennast kasuliku abivahendina infotalletuses ja -otsingus, on integreeritud raamatukogusüsteeme tootvad ja müüvad firmad lülitanud nendesse võimaluse koostada ja kasutada masinloetavat tesaurust. Käsiraamatus "Library Systems in Europe: a Directory & Guide" [10] on neid loetletud üheksa: ADLIB, BIBIS, BIS, BOOK PLUS,

CAIRS-LMS, ORACLE LIBRARIES, TECHLIB PLUS, TINLIB ja UNICORN. Nimetatud teatmik käsitleb ainult Euroopa Liidu maades (seisuga 1994) levinud kommertssüsteeme. Raamatus toodud lühiaandmete järgi on raske hinnata nende väärtusi ja puudusi, seepärast vaid paar märkust.

Tesaurus moodustab tavaliselt osa autoriseeritud andmete ohje (*authority control*) süsteemist, mis omakorda kuulub kataloogimismoodulisse. Keerulise struktuuriga tesaurusi integreeritud süsteemides enamasti koostada ei saa. Tavaline tulemus on standardse ülesehitusega tähestikuline esitus, mis eraldab deskriptorid ja sünonüümid ning näitab ära laiemad, kitsamad ja assotsieeruvad terminid. Teistest rohkem võimalusi pakub Saksamaa süsteem BIS, kus märksõnu saab täiendada notatsiooniga ning esitada tesaurust nii alfaabetilisel kui ka süstemaatilisel kujul. BISi tesaurusemooduli abil on koostatud suuri tesaurusi, mis kõlbavad ka trükis avaldamiseks.

Kaks- või mitmekeelset tesaurust saab hallata süsteemides ADLIB, BIBIS ja TECHLIB PLUS.

Peale märksõnade ükshaaval sisestamise on vahel võimalik ümber laadida masinloetavas vormis terviklikke tesaurusi, mille struktuur vastab selles süsteemis kasutatule.

Tesauruse abi märksõnastamiseks ja infootsinguks on eri süsteemides erineval tasemel. Mõni on suunatud rohkem raamatukoguhoidjale, teine lugejale. Päris hea tundub olevat ORACLE LIBRARIES. Selle tesaurust saab vaadata ja sealt sõnu noppida nii indekseerimise kui ka infootsingu ajal. Otsingut saab tesauruse abil laiendada või kitsendada. Märksõnu on võimalik seada vastavusse liigiindeksitega. ORACLE'i üht versiooni on demonstreeritud ka Eestis.

Tesauruse ja arvuti vahekorrast rääkides ei saa mainimata jätta tesauruste automaatse konstrueerimise katseid. Need arendavad seda suunda, mida infotesauruse idees nägi Luhn. Et inimhõimust arvutil modelleerida (veel?) ei osata, põhineb tesauruse automaatne konstrueerimine eelkõige statistilistel meetoditel. Tesauruseks organiseeritav sõnavara kogutakse mingist tekstikorpusest, näiteks tekstiandmebaasist. Sõnad rühmitatakse võtmesõnade klastritesse (alamhulkadeks), arvestades nende koos esinemise sagedust. Eelduseks võetakse, et sagedamini koos esinevad sõnad on tähenduslikult seotud. Täpsemaid semantilisi suhteid statis-

tiliste meetoditega muidugi tuvastada ei saa. Kui lisaks rakendada keele automaattöötluks mõeldud lingvistilisi meetodeid, on võimalik peale üksiksõnade tekstist eraldada grammatilisi fraase ja kokku viia samatüvelisi sõnu.

Automaatselt genereeritud tesauruse kasutuskõlblikkus sõltub sellest, milline tekstikorpused on võetud terminite allikaks. Kui selleks on mingi kindla andmebaasi tekst, ei pruugi saadud klastrid sobida otsinguks teistes andmebaasides, isegi kui need on samast ainevallast. Masinloetavate sõnaraamatute, eriti seletussõnaraamatute põhjal konstrueeritud tesaurusi saab rakendada erinevates keskkondades ja nad peegeldavad hästi semantilisi suhteid, kuid puudu jääb erialaterminoloogiast [9, 263].

Statistiliselt moodustatud sõnaklastreid saab kasutada automaatselt märksõnastamiseks: kui programm on uue teaviku tekstist esinemissageduse järgi välja valinud võtmesõna, otsib ta üles, millisesse klastrisse see kuulub, ning lisab teavikule märksõnadena ka teised klastri sõnad. Teine võimalus on kasutada klastreid ainult otsingufaasis: kui päring sisaldab mõne klastrisse kuuluva sõna, laiendatakse otsingut automaatselt kõigile klastri sõnadele.

Automaatselt konstrueeritud tesaurusi on rakendatud vaid vähestes infosüsteemides. Raamatukogukataloogides pole seda üldse tehtud, sest kataloogikirjetes esile tuleva sõnavara statistiline töötlus ei annaks mõtet tulemust.

Lõpetuseks. Infotesaurusega seondub palju huvitavaid probleeme, mida selles ülevaates on vaid riivamisi käsitletud. Palju jäi hoopis puudutamata. Tundub, et tesaurused kui infokeeled pole end sugugi ammendanud ja nende loomine jätkub, kuigi see töö nõuab palju aega ning oskajaid inimesi. Aga kui informatsioon on kallid, miks peaksid selle töötlemise vahendid siis odavad olema?

Kirjandus

1. Aitchison, J., Gilchrist, A. Thesaurus construction : a practical manual. 2nd ed. London, 1987. 173 p.
2. Bates, M. J. Subject access in online catalogs : a design model. // Journal of the American Society for Information Science. 37 (1986) 6, p. 357-376.

3. Chan, L. M., Pollard, R. Thesauri used in online databases : an analytical guide. New York, 1988. 268 p.
4. Classification systems and thesauri 1950-1982. Frankfurt, 1982. 143 p. (International classification and indexing bibliography; 1).
5. ISO 2788 : Documentation. Guidelines for the establishment and development of monolingual thesauri. 2nd ed. 1986. 32 p.
6. ISO 5964 : Documentation. Guidelines for the establishment and development of multilingual thesauri. 1985. 61 p.
7. Krooks, D. A., Lancaster, F. W. The evolution of guidelines for thesaurus construction. // Libri. 43 (1993) 4, p. 326-342.
8. Lambert, N. Of thesauri and computers : reflections on the need for thesauri. // Searcher. 3 (1995) 8, p. 18-23.
9. Lancaster, F. W., Warner, A. J. Information retrieval today. Arlington, 1993. 341 p.
10. Library systems in Europe : a directory & guide. London, 1994. 401 p.
11. Milstead, J. L. Thesaurus management software. // Encyclopedia of library and information science. Vol. 51 (Suppl. 14). New York, 1993. P 389-407.
12. Märksõnaliigitus : metoodiline materjal / koost. H. Masing. 2. tr. Tallinn, 1987. 36 lk.
13. The Penguin Roget's thesaurus of English words and phrases. New York, 1985. 776 p.
14. SFS 5471 : Suomenkielisen tesauruksen laatimis- ja ylläpito-ohjeet. // Tietohuolto. Helsinki, 1994. S. 171-179.
15. Thesaurus guide : analytical directory of selected vocabularies for information retrieval. Amsterdam, 1985. 749 p.
16. Vickery, B. C. Thesaurus - a new word in documentation. // Journal of documentation. 16 (1960) 4, p. 181-189.
17. Õim, A. See põnev, tüütu, vaevaline sõnaraamatutöö : [M. Jõgi vestlus A. Õimuga]. // Kultuurileht (1995) 29. sept., lk. 12-13.

1997

LIBRARIES in the INFORMATION SOCIETY



International conference on library automation in Central and Eastern Europe

Conference Proceedings
Budapest 10-13 April 1996

Edited by **Monika Segbert**
Katarina Steinwachs
Peter Burnett

*Organized by the Regional Library Program of the Open
Society and DG XIII/E-4 "Libraries and Electronic Publishing"
of the European Commission*

European Commission, DG XIII / E-4

EUR 17551 EN

CONTENTS

I. INTRODUCTORY SESSION

INTRODUCTION TO THE CONFERENCE..... 3
Peter Burnett, Chairman, Regional Library Program Board, Open Society Institute

THE LIBRARIES PROGRAMME OF THE EUROPEAN COMMISSION AND
LIBRARY CO-OPERATION WITH CENTRAL AND EASTERN EUROPE..... 7
Ariane Iljon, European Commission DGXIII/E-4, Luxembourg

II. AUTOMATION

LIBRARY AUTOMATION IN CENTRAL AND EASTERN EUROPE:
PROGRESS AND PROSPECTS..... 13
Christine L. Borgman, Department of Library and Information Science, University of California, USA

HOW TO SELECT A LIBRARY SYSTEM..... 23
Martin Svoboda, National Library, Czech Republic

CO-OPERATION AND LIBRARY AUTOMATION IN ALBANIA..... 29
Luljeta Buza, National Library, Albania

AUTOMATION PROCESSES IN THE NATIONAL LIBRARY OF POLAND
THE NATIONAL CONTEXT..... 33
Jadwiga Sadowska, Ewa Krysiak, The National Library, Poland

ACTIVITIES OF THE AUTOMATION COMMITTEE OF THE BELARUSIAN LIBRARY
ASSOCIATION..... 39
Petr Lapo, Jakub Kolas Central Science Library, Belarus

THE ROLE OF LIBRARY AUTOMATION IN NATIONAL ECONOMIC
DEVELOPMENT..... 45
Robert M. Hayes, University of California, USA

III. NATIONAL BIBLIOGRAPHIES

THE BRITISH LIBRARY NATIONAL BIBLIOGRAPHIC SERVICE AND ITS ROLE IN
THE DISSEMINATION OF THE UK NATIONAL BIBLIOGRAPHIC DATABASE..... 59
James Elliot, British Library, United Kingdom

THE HUNGARIAN NATIONAL BIBLIOGRAPHY..... 67
Susanne Berke, The National Széchényi Library, Hungary

THE SLOVENIAN NATIONAL BIBLIOGRAPHY ON CD-ROM..... 73
Maja Zumer, National and University Library, Slovenia

IV DOCUMENT DELIVERY

DOCUMENT DELIVERY: ACCESS VERSUS HOLDINGS.....	79
David Baker, University of East Anglia Library, United Kingdom	
DECOMATE: A NEW LOOK AT ELECTRONIC DOCUMENT DELIVERY A COMPARATIVE OVERVIEW OF ELECTRONIC DOCUMENT DELIVERY SYSTEMS.....	89
Joost Dijkstra, Tilburg University Library, The Netherlands	
FASTDOC AND RELATED EUROPEAN PROJECTS ON ELECTRONIC DOCUMENT ORDERING AND DELIVERY.....	99
Catherine Synellis, University of Patras, Central Library, Greece	

V BIBLIOGRAPHIC STANDARDS

ELECTRONIC COMMUNICATION BETWEEN LIBRARIES - STANDARDS, PROTOCOLS AND EUROPEAN INITIATIVES.....	113
Gordon Pedersen, Fischer & Lorenz, Denmark	
UNICODE / ISO 10646 AND THE CHASE PROJECT.....	119
Anthony Brickell, British Library, United Kingdom	
EBSCO'S EDI / INTERFACE PARTNERS.....	127
Hanna March, EBSCO, United Kingdom	
BIBLIOGRAPHIC STANDARDS IN A SHARED CATALOGUING SYSTEM: (INTER)NATIONAL CO-OPERATION, INTERESTS, NEEDS.....	133
Zlata Dimec, National and University Library, Slovenia	
INTERNATIONAL STANDARDS AND THE CREATION OF INTERNATIONAL NAME AUTHORITY FILES: WHAT IS MISSING?.....	139
Mirna Willer, National and University Library, Croatia	
THE IMPACT OF UNIMARC ON BIBLIOGRAPHIC WORK IN LITHUANIA.....	147
Elena Maceviciute, Vilnius University, Lithuania	
Liubovj Buckiene, Lithuanian Martynos Mažvydas National Library, Lithuania	
APPLICATION OF STANDARD GENERAL MARKUP LANGUAGE (SGML) IN A LIBRARY ENVIRONMENT.....	153
Vitaly N. Tresorukov, Artyom A. Terekhov, Vadim A. Sobolev, All-Russia State Library for Foreign Literature, Russia	
THESAURUS FOR SUBJECT SEARCHING AND INDEXING AS PART OF AN INTEGRATED LIBRARY SYSTEM.....	157
Sirje Nilbe, Tartu University Library, Estonia	
RUSSIAN STATE LIBRARY FORMATS FOR THE PRESENTATION OF DATA IN MACHINE-READABLE FORM.....	161
N.V. Sarycheva, Russian State Library, Russia	

THESAURUS FOR SUBJECT SEARCHING AND INDEXING AS PART OF AN INTEGRATED LIBRARY SYSTEM

Sirje Nilbe
Tartu University Library, Estonia

Introduction

Problems of subject access attract much attention amongst researchers and designers of library systems. However, one has to concede that the practical results are still far from satisfactory. In general, libraries have continued to use the same controlled vocabularies and indexing methods as for card catalogues, indexed by pre-co-ordinated subject headings like Library of Congress Subject Headings (LCSH). There are two essential reasons for this:

- the design of new vocabularies more suitable for post-co-ordinated indexing and searching is expensive and time-consuming;
- re-indexing of all library materials for retrospective cataloguing is considered to be very time-consuming.

Libraries are always and everywhere short of staff and funds. Most efforts to improve subject access are therefore made post factum (by automated processing of text of records and by developing user interfaces).

The large academic libraries in Estonia have arranged their subject card catalogues systematically by UDC for over 40 years. It is considered reasonable to also add UDC numbers to electronic records because there is a need for a standardised and language independent subject access on a national and international scale. UDC tables in Estonian will be published this year. UDC numbers have some good searching facilities for the experienced user, who can broaden or narrow the query with ease. But for the inexperienced user the numbers are meaningless. It is clear that there is not much enthusiasm about UDC number searching even amongst librarians.¹ That is why the main facility for subject access for end-users should be verbal.

Software

Tartu University Library uses INGRID, an integrated library system developed in-house. It is based on the INFORMIX database management system and runs on a UNIX platform. Up to now the following modules have been implemented: acquisition, serials control, cataloguing, OPAC. For vocabulary control and subject searching there is a thesaurus module.

Thesaurus software can be more or less complicated. INGRID poses minimal requirements for thesaurus software. It is possible to create conventional relationships as USE, BT, NT and RT. Reciprocal relationships are formed automatically when input is made. J. L. Milstead² declares that automatic reciprocation of relationships is the fundamental criterion for considering a software package to support thesaurus management. In addition to fields for referred terms, there are fields for English equivalents, scope notes and broad subject domain names in a term

¹ Buxton, A. B. Computer searching of UDC numbers. In: *Journal of Documentation*, vol. 46, no. 3, pp. 193-217.

² Milstead, J. L. Thesaurus management software. In: *Encyclopaedia of library and information science*, vol. 51. New York: Marcel Dekker 1993, pp. 389-407

record. Term validation capabilities are good (it is not possible to create a new record for an existing term or refer to a term not yet included in the thesaurus). However, relationship validation is not realised automatically. For example, errors like having the same term both on the BT field and RT field are not automatically avoided. The main display of the thesaurus is alphabetical. Any string of terms can be retrieved using left, right and middle truncation. Right truncation is implicit while searching the thesaurus database (not while searching for references in the catalogue). Rotated listings formed this way provide an entry not only for every word but for every part of the word. This is very important because there is a large number of compound words and derivative suffixes in Estonian.

For systematic grouping of terms there are the subject domains like chemistry, economy, medicine. Further classification is implied by BT/NT relationships only.

The use of the thesaurus for indexing is similar to that for searching. The user opens a new window and displays an alphabetical listing of terms. Then he/she can navigate through the thesaurus using relationships between terms.

While indexing a document it is only possible to add descriptors included in the thesaurus by selecting terms from the list. Uncontrolled terms cannot be put in. A non-preferred term is automatically replaced by the preferred term when used in indexing or searching. While searching for references, selecting from the list is not compulsory. One can also input the search terms into a search form by using truncation and the logical operators AND, OR, NOT

It is possible to display a print preview of any term with its relationships and other appropriate information. Printing the thesaurus or part of it is a facility available to staff.

Terms in the thesaurus are both in Estonian and in English, the former as the dominant language. The indexing language is Estonian. English terms are linked to Estonian ones and are searchable only from the OPAC if English is chosen as the dialogue language.

Due to the very dynamic nature of our thesaurus, the full integration of thesaurus maintenance and database updating is absolutely essential. A change in a thesaurus term causes a global change to the database: all documents indexed with the old version are changed to the new version of the term. Deleting a term from the thesaurus is only possible if there are no documents linked to this term, i.e. corresponding records must be edited first. This is time-consuming but improves the quality of the database.

It seems that the facilities for vocabulary control offered by the thesaurus module of INGRID are no more limited than those offered by most corresponding modules of commercially available library systems.³

Thesaurus contents

The difficulties connected with thesaurus contents and structure are more serious than software problems. There is no complete thesaurus in Estonian except one for indexing literature on Finno-Ugristics for the international reference database URBIS. The general Estonian thesaurus for libraries is being developed in the National Library, but it was in its initial stages when the University Library of Tartu began indexing for the on-line catalogue in November 1994, and it

³ Library systems in Europe: a directory & guide. London: TFPL Publishing 1994.

is still far from being completed. Building the thesaurus and indexing documents must therefore be done at the same time.

The international standards ISO 2788, ISO 5964, ISO 5963, and well-known handbooks⁴ are used as guidelines. But there are some specific factors affecting the indexing and thesaurus construction environment in Tartu university library. These are the very broad subject and language coverage of collections and the lack of fixed scientific terminology in Estonian.

Tartu University Library is an old and large library, celebrating its 200th anniversary in 2002. The collections of the library (3,7 million items) include research monographs, textbooks, fiction and other non-scientific materials, pictures and manuscripts etc. Subject and language coverage comprise exact sciences, natural sciences, social sciences, humanities, in Estonian, English, German, Russian, Finnish, Swedish etc. It is very difficult to build a unified conceptual system for indexing this universe of knowledge.

The thesaurus standards do not work very well in a universal library environment, as they are oriented towards the development of thesauri for specialised bibliographic databases. The required pre-co-ordination level is higher for a library catalogue than for a specialised database. There are cases where the strong semantic restrictions for the establishment of BT/NT relationships may hinder the effectiveness of thesaurus navigation. There appears to be an increase in end-user oriented thesauri in libraries and there is an imperative need for guidelines and handbooks related to this area.

Another problem connected to the content of the thesaurus is less library dependent: great changes are taking place in the conceptual basis of most subject fields in present-day Estonia. Regarding the use of language it results in extensive synonymy and vagueness of meanings. Fortunately terminology as a field of applied linguistics has quite a long tradition in Estonia. Terminologists and subject specialists are working intensively on vocabularies in several subject fields. It is to be hoped that Estonian terminology will soon be ready and that the thesaurus will be a useful tool for indexing and searching in on-line library catalogues.

⁴ Aitchison, J., Gilchrist, A. Thesaurus construction: a practical manual. 2nd ed. London: Aslib 1987 and: Lancaster, F. W. Vocabulary control for information retrieval. 2nd ed. Arlington: Information Resources Press 1986, 270 p.

Raamatukogu

- * Ellen Niit, Anne Rande, Andres Jaaksoo jt. lapsest lugejana,
- * Vestlus V. Beekmaniga lastekirjanduse tõlkimisest,
- * Astrid Lindgreni uus lugu Pipist,
- * Aino Pervik almanahhist "Nukits" jm.

2 / 1997



SISUKORD

■ AVAVEERG

Laps ja raamatukogu

Ellen Niit 4

■ LASTERAAMATUKOGU

Iga lugeja sünnib lapsena

Anne Rande 5

Eesti maaraamatukogude lapsi kummitab lugemisnälg

Andres Jaaksoo 8

Väikese juubeli aasta

Ädu Neemre 10

Laste arvel kokku hoida ei tasu: intervjuu Rapla KRK raamatukoguhoidja Leili Ojaga.

12

Karupoeg Puhh ja taoism

Mait Talts 14

Astrid Lindgreni uus lugu Pipist

18

■ JUTUTUBA

Praegu on oluline säilitada Eesti laste lugemishuvi: intervjuu V Beekmaniga

Maire Liivamets 20

■ RAAMATUKOGUNDUSE UUDISKIRJANDUS

23

Eesti lasteraamatukogud 1996. aastal

27

Eesti rahvaraamatukogud 1996. aastal

30

■ SAAGEM TUTTAVAKS

Kuressaare Gümnaasiumi raamatukogu

Urve Aedma 31

■ RAAMATULUGU

TÜRI haruldasi raamatuid: Litograafia tulek raamatukunsti

Tiina Nurk 32

■ EESTI MÖTTELUGU

Võitlus kestab edasi

Hilve Rebane 34

■ ARVUSTUS

"Nukitsast" ja missioonist

Aino Pervik 36

■ LUGEJA ARVAB

Meist enestest

Anne Rande 37

■ PÄEVAKORRAL

Märksõnastamine Tartu Ülikooli Raamatukogus

Sirje Nilbe 38

■ SÕNUMID

39

Õnnitleme: Mari Kalvik – 60

45

■ SUMMARY

46

Mina vaidlen vähemalt esialgu sellele vastu. Välimus on tihtigi otsustavalt tähtis, pean siin lisaks riitusele silmas ka naeratust ja lahkete silma-vaadet. Sest võta või jäta, esimene mulje inimesest avaldub just eelpool-õeldut silmas pidades.

5. versioon: Või on see tingitud vaesusest meis ja meie ümber? Aga kas me ikka oleme kõige vaesemad? Arvan et see pole vaesus, kui näiteks üks suhteliselt tuntud raamatukogu-tädi asjalikul töökoosolekul tagu-mises reas oma kotist vaikselt saia nosib, ümberringi piiludes, ega keegi juhuslikult ei näe.

Niipalju siis minu oletustest.

Eva Lille artikkel "Raamatukogus" nr. 4, 1995 pahandas meist paljusid. Mina olin väga pahane ja teatud mõttes olen seda veelgi, kuna selles kritiseeriti kolleege Pärnust, kes tegelikult on väga kenad, toredad, elurõõmsad inimesed. Aga andkem tõe au, eks pr. Lille kirjutises on ka tera tõtt – on ju need tema kui kõr-valtvaataja visioonid Eesti raamatu-kogutädidest. Muide, ka see väljend ... tädi on ajast ja arust. Küll oleks kena, kui selle saaks ära keelata!

Kus aga tegelikult peituvad ülal-toodud arvamustele-oletustele vastu-sed? Ja kas neid ongi? Ei tea! Tean aga seda, et juba 7-aastase plikana oli mul kaks unistust: kas minust saab Helmi Puur (loe: baleriin) või raama-tukoguhoidja. Kuna esimese soovi täitumine oli tol ajal maakohas või-matu, siis jäi alles teine. Raamatu-koguhoidjaks tahtsin seepärast, et sain istuda laua taga, tähtsa olekuga, suhelda inimestega ning täita nende lugejakaarte. Uhke oli tunda end suure raamaturiigi valitsejana, sain ka ise lugeda kõiki raamatuid, isegi "Dekameroni"! Mulle meeldis selga panna ilusaid riideid, et lugejad saaksid imetleda, kui kena ma välja näen. Mis sest, et kõik olid ema vanadest tehtud. Loomulikult oli minu suureks eeskujuks kohalik raamatukoguhoidja, punase juukse-pahmaka ja tedretähnidega, keda püüdsin igati matkida, kelle kohta aga küllarahvas rääkis: "Näe, jälle ta läheb, kotid raamatuid täis, küla peale jagama! On alles imelik!" Nüüd kuulun juba pikka aega ise nende imelike hulka. Ja ikka veel meeldib!

Autori tekst muutmata

MÄRKSONASTAMINE TARTU ÜLIKOOLI RAAMATUKOGUS

SIRJE NILBE

pearaamatukoguhoidja

Tartu Ülikooli Raamatukogus on märksõnastamisega tegeldud juba kahekümnendate aastate algusest, mil F Puksoo initsiatiivil hakati koostama käsikogude sõnastik-katalooge. Need said aluseks praegustele **vanema kirjanduse märksõnakataloogidele**, mis hõlmavad raamatuid ilmumisaastaga kuni 1944. Eestikeelse kirjanduse kataloogis on praegu 1100 ja võõr-keelse kirjanduse kataloogis 3600 märksõnarubriiki. Hulk aastaid on neid katalooge korraldanud Lea Trikkant, kes on lõpule jõudmas põhjaliku redigeerimistööga. Tema koostatud on ka nii lugejaid kui märksõnastajaid abistavad tähestikulised ja süstemaatilised märksõnaloendid.

Vanema kirjanduse märksõna-kataloogid täienevad pidevalt, sest raamatukokku hangitakse jätkuvalt enne 1945. aastat ilmunud kirjandust. Lisa tuleb ka rekataloogitud raamatute hulgast. Eestikeelset kirjandust märksõnastab Lea Trikkant üksi, võõrkeelset ka teised liigitajad. Eesmärgiks on peegeldada kogu vanem raamatuvara ühtsetes märksõnakataloogides.

Pikka aega on raamatukogus märksõnastatud Eesti-ainelist võõrkeelset kirjandust. Selle töö tulemused peegelduvad **vanema Estica sõnastikkataloogis** (teosed ilmumisaastaga kuni 1944) ja **uuema Estica märksõnakataloogis**. Viimast koostab Ruth Michelson. *Estica* märksõnastik on viimistlusjärgus, sisaldades praegu üle 700 rubriigi. *Estica* märksõnakataloog täieneb nii saabuva

kui ka varemkomplekteeritud sõjajärgse kirjanduse põhjal.

Teatmekirjanduse märksõnakataloogi, mis peegeldab teatmeteoseid 1948. aastast alates, hool-dab Ljudmila Dubjeva. Teavikuid märksõnastavad erialaliigitajad. Praegu on kataloogis 1400 rubriiki.

Peale nimetatute on märksõna-või temaatilisi katalooge veel haru-kogudes ning käsikirjade ja harul-daste raamatute osakonnas (käsi-kirjad, fotod). Tartu Ülikooli bib-liograafia ja Eesti ajaloo bibliograa-fia töökartoteegid, mis asuvad teadusbibliograafia osakonnas, on samuti suures osas märksõnalised.

Seni on jutt käinud kaartkata-loogidest. Tartu Ülikooli Raamatu-kogul on juba küllalt suured kogemused ka elektrooniliste andme-baaside märksõnastamisega. Oluli-sim neist on **elektronkataloog INGRID**, mida alustati 1994. aasta novembris. Sinna sisestatakse ladina kirjas väljaanded ilmumis-aastast 1990. Märksõnastamist elektronkataloogis on juhendanud allakirjutanu.

INGRIDI tesauruses on hetkel ligi 12 000 deskriptorit ja viite-terminit, mis jagunevad 49 aine-valdkonna vahel. Märksõnastatud on 28 000 nimetust raamatuid, jadaväljaandeid ja kaarte. Iga liigi-taja on lisanud oma valdkonda uusi märksõnu, leidnud inglis-keelsed vasted, loonud semantili-sed seosed teiste märksõnadega. Käesoleva aasta algusest alustasid tesauruse süstemaatilist läbivaata-mist ja toimetamist Mare Onga (humanitaarteadused) ja Tiit

larkpea (reaalteadused). Paljude valdkondade jaoks ei leidu raamatukogus siiski erialainimesi, kes vabalt orienteeruksid kaasaja teaduses ja selle mõistestikus. Üksiküsimustes on konsulteeritud ülikooli õppejõudude ja teaduritega, kuid nende pikemaajalisem kaasamine vajaks lisarahastamist.

Lugejad saavad kasutada INGRIDi tesaurust kataloogiotsingu käigus abimenüüna.

Ülikooli raamatukogu osaleb rahvusvahelises **soome-ugri keeleteaduse bibliograafia** projektis URBIS. Tema ülesanne on koguda andmeid endise Nõukogude Liidu aladel ilmuva fennougristikaalase kirjanduse kohta (ilumisaastast 1988). Kogutud kirjed sisestatakse andmebaasi ja märksõnastatakse URBISe märksõnastiku järgi. See on tesauruse vormis käsikirjaline märksõnastik, millest on ühilduvad variandid eesti, soome, ungari, inglise, saksa ja vene keeles. Tartu kirjed märksõnastatakse eesti, inglise ja vene keeles. URBISe andmebaasi, milles praegu on ligikaudu 2000 kirjet, koostab Mare Onga.

Teadusbibliograafia osakonnas sisestatakse arvutiandmebaasi ka Eesti ajaloo **bibliograafia** (aastad 1918-1944) kartoteeki. Märksõnastamisel kasutatakse kartoteegi rubriike ja vabu märksõnu. Andmebaasis on u. 9000 kirjet. Ajaloo-bibliograafiat koostab neli inimest Kersti Taali juhtimisel.

Seega on märksõnastustöö ülikooli raamatukogus küllalt mitme-palgeline. Koostatakse ja kasutatakse mitmeid eri märksõnastikke ja erinevaid märksõnastuse metoodikaid lähtudes andmekogu sisust ja kujunenud traditsioonist. See on paratamatu. Sellist metoodikat, mis annaks igas olukorras ühtviisi häid tulemusi, pole tõenäoliselt olemas, samuti pole ka võimalik koostada kõikehõlmavat märksõnastikku.

23. jaanuaril toimus Tartu Ülikooli Raamatukogus järjekordne **FRIEDRICH PUKSOO PÄEV**

Raamatukogu vastremonditud (kuigi veel vajaliku tehnikaga sisustamata) saali koguneti õnnilema F. Puksoo preemia laureaate ja kuulama ettekandeid eesti raamatu ajaloost. Üritust avades rõhutas TÜRi direktor Peeter Olesk, et 1997. aastal on oluline Eesti sundeksemplariseaduse vastuvõtmine: "Riigikogugi peaks lõpuks taipama kui oluline on OMA – EESTI RAAMATU!" Ta soovis raamatu- ja raamatukogu-uurijale kannatlikkust, huumorimeelt ja avastusi nii inimeste kui raamatute vallas.

Friedrich Puksoo preemia, viimasel aastal trükitud ilmunud prima raamatuloo, raamatuteaduse, raamatukogunduse ja bibliograafiaalase kirjutise või bibliograafia autorile määratava preemia saamiseks oli esitatud 10 tööd. Parimaks tunnistati **Kaljo-Olev Veskimägi raamat "Nõukogude unelaadne elu: Tsen-suur Eesti NSV-s ja tema pere-mehed"** (Tallinn, 1996). Žürii esinaine Mare Lott põhjendas preemiat üle andes otsust nii: "Uudsel teemal valminud uurimus toob teaduslikku ringlusesse ulatusliku seni kasutamata arhiivimaterjali, analüüsib faktiandmetele tuginedes tsensuuri toimetehhanisme ja vaimuelu ohjamise nõukogulikku süsteemi." Lisaks preemiale sai K.-O. Veskimägi teise nummerdatud eksemplari ERÜ poolt välja antud sajaeksemplarilisest faksiimiletrükist F. Puksoo "Raamatuharastusest ja Jaan Roosist kui bibliofiilist" (Väljaande esimene eksemplar kingiti TÜRile.) Esitatud töödest **märgiti ära Arvo Teringu stiilne uurimus "Descartes ja tema ideede jõudmine Baltimaile 17. saj. ja 18. saj. algul"** mis käsitleb filosoofia ajaloo ühte sõlmpunkti – uusaaja filosoofia kujundaja Rene Descartes'i retseptsiooni Lääne-Euroopas ja Baltimaades. Autorit auhinnati eelnimetatud faksiimiletrüki kolmanda eksemplari-ga.

Teaduspäeva ettekanded olid Eesti raamatust ja raamatukogundusest. **Marje Aasmetsa käsitus "Friedrich Puksoo – eesti raamatukogunduse rajajaid"** andis F. Puksoo 1930. aastate II poolel avaldatud artiklite põhjal ülevaate F. Puksoo huvitava-

matest seisukohtadest teadusraamatukogude ülesannete kohta ning raamatukogude rollist teadlaste uurimistöös. Pikemalt peatus ta Puksoo tööl kataloogimisjuhiste koostamisel, märksõnastamise algatamisel, oskussõnavara loomisel, raamatukogundushariduse arendamisel jm.

Huvitav oli **Tiiu Reimo** ettekanne "**Baltica leide Briti Raamatukogus**" P. Oleski hinnangul on see töö teedrajava tähtsusega, sest nüüd, kui meil on võimalus käia uurimistöös tegemas maailma suuremates raamatukogudes, saame käsile võtta ülevaatlikuma ja sügavama töö eesti raamatu arenguloo uurimisel. Briti Akadeemia stipendiaadina viis nädalat Briti Raamatukogus töötanud T. Reimo tegi huvitavaid leide nii Eestis trükitud kui Eestiga seotud 16.-18. sajandi trükiste osas. Mõned ettekandes käsitletud intrigeerivad väljaanded: Tallinnas 1723. aastal trükitud 7 rootsikeelset pietistlikku raamatukest sisaldav konvoluut; vanim Eestit käsitlev teos Briti Raamatukogus, Strasbourgis 1571. aastal trükitud viieleheküljeline luuletus vene sõdalaste metsikustest Tallinnas 1571. aasta mihklipäeval; Westphalist pärit Christophorus Dudulaeusele omistatud raamat igavesest juudist Ahasveerusest, mis Euroopas 17. sajandil rahvaraamatuna mitmetes trükkides levis ja mida Briti Raamatukogus on koguni kolm väljaannet.

Kersti Pedak rääkis Tartu Ülikooli Raamatukogu komplekteerimis põhimõtetest aastatel 1919-1940, **Kiira Schmidt** saksa 17. sajandi raamatust Tartu Ülikooli Raamatukogus, **Sulo Lembinen** TÜRi fotokogust. **Peeter Oleski** "**Estica – mis see praegu on ja kuidas seda Tartus õpetada**" käsitles termineid rahvustrükis, *estica*, *estonica* (Kaja Noodla lisas ka termini *estonica*) ja nende sisulisi erisusi. Põhimõtteline probleem on, kuidas *estica* määratleda ja ammendavalt koguda. P. Oleski arvates tuleks seda teha sarnaselt soomlaste *fennicale*.

Konverentsi materjalid on saada-val RRi ja TÜRi raamatukogunduse teabetoas ning ETARi raamatukogunduse lugemissaalis.

Teaduspäev lõppes küünalde süütamisega F. Puksoo haua-l.

Ene Riet

TARTU ÜLIKOOLI RAAMATUKOGU TÖID

X



Tartu 1997

SISUKORD

Saateks	7
Foreword	9
K. Miil	
Uuring raamatukogu tegevuse tulemuslikkusest	11
<i>The survey of resultfulness of library performance</i> <i>at Tartu University Library. Summary</i>	21
Lisad / Appendices:	
Vorm 1–1. Küsitlusleht lugejate rahulolu kohta	23
Vorm 3–1. Raamatukogu materjalide kasutamine	25
Vorm 5–1. Küsitlusleht materjalide kättesaadavuse kohta	26
Vorm 13–1. Teatmeteeninduse statistika	27
Vorm 14–1. Küsitlusleht teatmeteeninduse taseme kohta	28
Vorm 15–1. Infootsingu hindamisleht	30
<i>Form 1–1. General satisfaction survey</i>	32
<i>Form 5–1. Materials availability survey</i>	34
<i>Form 14–1. Reference satisfaction survey</i>	35
<i>Form 15–1. Information search evaluation</i>	36
Teeninduspunktide lühendid tabelites	38
U. Tõnnov	
Lugeja probleemid on raamatukogu probleemid	39
<i>The problems of the user are the problems of the</i> <i>library. Summary</i>	58
E. Loorits, L. Dubjeva	
Teatmeteenindus Tartu Ülikooli Raamatukogus	61
<i>Reference services at Tartu University Library.</i> <i>Summary</i>	74
L. Dubjeva	
Kataloogid — konsultant — lugejad	77
<i>Catalogues — consultant — users. Summary</i>	86
M. Ilus	
Materjalide kättesaadavus ja lugejate rahulolu	89
<i>The accessibility of library materials and user</i> <i>satisfaction. Summary</i>	93

Ü. Must

Tagasiside kasutajalt kui tähtsaim stiimul info-teenistuse arendamisel	95
<i>Feedback from the users as the most important impetus in developing information services.</i>	
<i>Summary</i>	104

T Matsulevitš

Raamatukogu kui lugeja töökeskkond	106
<i>Library as the user's work environment. Summary</i>	112

★

M. Ermel

Raamatukogud infoajastul	113
<i>Libraries in the information era. Summary</i>	127

I. Saare

Tartu Ülikooli Raamatukogu elektronkataloog	129
<i>Electronic catalogue of Tartu University Library.</i>	
<i>Summary</i>	133

M. Meltsas

Kaasaegse infotehnoloogia rakendamisega seotud psühholoogilised probleemid	135
<i>Some psychological problems connected with the application of modern information technology. Summary</i>	143

M. Kümnik

Tartu Ülikool bibliograafias	144
<i>Tartu University in bibliographies. Summary</i>	165

S. Nilbe

Universaalse kümnendliigituse (UDK) rahvusvahelisest levikust	169
<i>The worldwide use of the Universal Decimal Classification. Summary</i>	182

M. Halbert

Tartu Ülikooli Raamatukogu automatiseerimise plaan 1995–1997	184
<i>Tartu University Library automation plan 1995–1997 Summary</i>	211

UNIVERSAALSE KÜMNENDLIIGITUSE (UDK) RAHVUSVAHELISEST LEVIKUST

Sirje Nilbe

Mõni aasta tagasi, kui Eesti raamatukogudes tunnetati, et paljutki peaks teistmoodi tegema, oli tunda kõhklust ka UDK-liigituse jätkamise suhtes. Suurte raamatukogude süstemaatilistest kataloogidest oli raske vajalikku kirjandust üles leida, käegakatsutavas läheduses terendasid elektronkataloogid, millest "keegi nägunii mitte ilalgi indeksi järgi otsima ei hakka". Liigitustabelid olid "nõukogude" ja vene keeles. Põhjamaade kolleegid pidasid UDK-d aegunud ja väljasurevaks nähtuseks, selle rahvusvaheline haldamine oli käest ära lastud jne. jne. Ometi tuli raamatukogunduslik konservatiivsus seekord kasuks ning UDK areng võttis konstruktiivse pöörde nii rahvusvahelises ulatuses kui ka Eestis, kus hakati ette valmistama esimest omakeelset UDK väljaannet [16].

1995. aasta sügisel tähistati universaalse kümnendliigituse 100. sünnipäeva. UDK loomislooc ja 75-aastase ajaloo kohta Eestis saab huviline lugeda M. Aasmetsa artiklit [1]. Käesoleva kirjutise eesmärgiks on lühidalt kirjeldada UDK rakendusvaldkondi ja levikut teistes riikides. Leitud andmed on küll katkendlikud ja ebaühtlased, sest liigituspraktika pole teema, mis teab kui sageli käsitlemist leiaks.

UDK rakendusala

Sünteesilise klassifikatsioonina on UDK struktuurilt väga paindlik. See annab võimaluse kasutada teda mitmeks otstarbeks, valides vajaliku detailsuse astme ning sünteesiviisi. Sõltumatus keelest ja süstemaatiline ülesehitus soodustavad rahvusvahelist infovahetust.

Indeksi võib teavikule omistada kahel viisil: 1) märkides selle teavikule endale, 2) lisades indeksi teaviku kohta koostatud kirjele. Esimesel juhul määratakse indekseerimisega peale teaviku sisu ja vormi sageli ka selle füüsiline asukoht, kui UDK on võetud paigutussüsteemiks.

Enamikus maailma raamatukogudes ongi liigitussüsteemid kogude paigutuse aluseks. UDK on levinud rohkem reaalteaduslikes ja tehnikaraamatukogudes. Suur universaalkogu on UDK järgi paigutatud näiteks Pompidou Keskuse raamatukogus Pariisis ning Tsukuba teaduskeskuse raamatukogunduse ja infoteaduse ülikoolis Jaapanis [22, lk. 103]. Paigutuseks tavaliselt UDK süsteemi lihtsustatakse.

Kui suurem osa kogusid paikneb kinnistes hoidlates, on raamatukogu teatmestu tähtsaks osaks **süstemaatiline kataloog**. UDK on ainuke üldtunnustatud liigitussüsteem, mis on loodud just süstemaatilise kataloogi vajadusi silmas pidades (teised on kujundatud eelkõige paigutusskeemidena). Bibliograafilistes kirjetes kasutatakse detailsemaid indekseid, kuid ka siin saab liigitustaset varieerida vastavalt kogude suurusele ja liigitatavate teavikute iseloomule. Süstemaatiline kataloog on eelkõige Euroopa nähtus.

Elektronkataloogile üle minnes on need raamatukogud, kes varem pidasid süstemaatilist kataloogi UDK järgi, tavaliselt jatkanud indeksi te lisamist kirjetele. Ehkki lugejad eelistavad konkreetsete päringute jaoks võimaluse korral märksõna- või võtmesõnaotsingut, on liigitus vajalik kataloogi sisust süstemaatilise pildi saamiseks. Sobiva tarkvara puhul on UDK elektronkataloogi "lehitsemiseks" parim vahend. UDK-otsingu põhi-probleem on selles, et tavaline lugeja ei valda UDK notatsiooni. Euroopas on välja arendatud mitmeid raamatukogusüsteeme, milles UDK indeksid ja märksõnad on omavahel automaatselt seotud, näiteks ETHICS (*Eidgenössische Technische Hochschule*, Zürich), CIS-DK (*Bereichsbibliothek Erziehungswissenschaften*, Gießen), ALIS (*Danmarks Tekniske Bibliotek*), HELECON (*Kauppakorkeakoulu*, Helsingi). 1992. aastal ilmunud LIBER-i teatmiku järgi [19] oli UDK liigitusstandardiks 18 Euroopa automatiseeritud raamatukoguvõrgus 49-st.

UDK-d on kasutatud **infoväljaannete** koostamisel kui materjali korrastamise vahendit. Siin tuleb kasuks UDK detailsus, sest artiklid, teadusaruanded, konverentsiettekanded jms. on tavaliselt kitsamatel teemadel kui monograafiad. Infoteenistustes on kergem minna kaasa skeemis tehtavate muudatustega, sest mahukat reklassifitseerimist nagu raamatukogus need kaasa ei too: regulaarselt ilmuva infoväljaande järgmises numbris kasutatakse lihtsalt uuendatud skeemi. UDK järgi on süstematiseeritud näiteks sellised infoväljaanded, nagu "Key to economic sciences" (Haag, 1978–), "Bibliographia cartographica" (München, 1974–) ja "Photographic abstracts" (Brentwood 1921–). Mitteperioodi-

listest bibliograafiaväljaannetest võib nimetada A J Walfordi "Guide to reference material", sama autori "Guide to current British journals in the humanities and the social sciences" ning *British Library* väljaannet "Current British journals" [22, lk. 105]

Trükitud infoväljaandeid on üha sagedamini asendamas **sidusandmebaasid**. Nendes kasutatakse indekseerimiseks ja infootsinguks tavaliselt deskriptoreid, kuid vahel ka UDK indekseid. 1985. aasta teatmikes leidub 39 Austraalias, Austrias, Hispaanias, Hollandis, Kanadas, Saksamaal, Soomes, Suurbritannias, Šveitsis, Taanis ja USA-s toodetud andmebaasi, mille otsikeeleks on UDK. Hõlmatavad alad on ehitus, mesindus, metsandus, ohutustehnika, glatsioloogia, meteoroloogia, okeanoloogia, hüdroloogia, autotööstus, majandusteadus, juhtimine, väliskaubandus jm. [36, lk. 15]. Praeguseks on andmebaaside arv nii tohutult kasvanud, et keegi neid kõiki enam registreerida ega kirjeldada ei jõua. Võib arvata, et UDK levik andmebaasides pole laienenud, sest tesaurust peetakse erialainfo otsingul efektiivsemaks vahendiks.

Laialt on levinud **normdokumentide** märgendamine UDK indeksiga, seda eelkõige ISO-ga (*International Organization for Standardization*) ühinenud maades, aga ka mujal. 1985. aastal oli neid riike 47 (tolleaegse poliitilise kaardi järgi), lisaks kasutavad UDK-d oma standardites regionaalsed ja rahvusvahelised organisatsioonid. Hõlmatud on kõik Euroopa riigid, v.a. Prantsusmaa ja Iirimaa, ning teiselt poolt kõik tähtsamad tööstusmaad, v.a. Kanada ja USA. UDK ise oli 1985. aastal rahvusliku standardina aktsepteeritud viies riigis: Suurbritannias, Austrias, Hispaanias, Tšehhoslovakkias ja Ungaris [13].

Küllalt levinud on UDK kasutamine **rahvusbibliograafiates**. UNESCO on seoses UBC (*Universal Bibliographic Control*) programmiga soovitanud struktureerida rahvusbibliograafia väljaanded süstemaatiliselt mõne üldlevinud liigitusskeemi järgi. Enamasti on selleks saanud DDC (*Dewey Decimal Classification*). 1993. aasta andmetel kasutatakse UDK-d rahvusbibliograafias 21 riigis [22, lk. 105]. See arv pole kindlasti enam täpne, sest arvesse pole võetud kõiki NSV Liidu ja Jugoslaavia lagunemisel tekkinud riike, kelle kohta andmed pole veel levinud. (Kasvõi näiteks Eesti on nüüdseks lisandunud.) Euroopa maadest on nimetada Belgia, Bulgaaria, Hispaania, Poola, Portugal, Prantsusmaa, Rumeenia, Saksamaa, Slovakkia, Soome, Tšehhi, Ungari. Nende riikide raamatutoodang moodustab umbes 50% kogu Euroopa toodangust. Ladina-Ameerika maadest kasutatakse UDK-d Argentiinas ja Ecuadoris ning regionaalbibliograafias "Notes bibliographiques

Caräibes" Aafrika riikidest on mitmel pool nimetatud Alžeeria, Benin, Madagaskar, Maroko, Sair, Senegal ja Tuneesia [23, 32].

Rahvusbibliograafiates leiduvad UDK indeksid on tavaliselt väga üldised. See ei häirinud perioodiliselt ilmuvates trükiväljaannetes, kuid elektronandmebaasides (magnetkettad, CD-ROM-id), kus andmed pidevalt kumuleeruvad, ei võimalda liiga üldine liigitus enam tulemuslikku otsingut.

UDK uusim rakendusvaldkond on **interneti** inforessursside temaatiline korrastamine. Mitmeid näiteid selle kohta leiab Suurbritannia infoserveritest. *NISS Information Gateway's* on ressursid korrastatud hierarhiliselt, kasutades UDK eri tasemeid (<http://www.niss.ac.uk/subject/index.html>). *The BUBL (Bulletin Board for Libraries) Information Service* pakub otsingu alustamiseks UDK süsteemi järgi reastatud lineaarse teemaloendi (<http://www.bubl.bath.ac.uk/BUBL/Tree.html>). Eelmisega sarnane, s.o. vaid üldisi UDK liike sisaldav on *University of Wales Aberystwyth* otsilehekülg (<http://www.aber.ac.uk/~infolib/classified.html>). Infoteenistus SOS!G võimaldab UDK järgi otsida sotsiaalteaduslikku infot (<http://sosig.esrc.bris.ac.uk/Subjects/udc-list.html>). Seda laadi alguspunkte leidub internetis teisigi, kuigi mitte palju.

UDK eri maades

Kõige rohkem on leida andmeid Euroopa riikide kohta.

Suurbritannias haldab UDK-d *British Standards Institution* (BSI). Täielikku väljaannet inglise keeles hakati publitseerima 1943. aastal. 1980. aastaks oli ilmunud vähemalt 107 vihikut [9, lk. 99], aga nende seas pole näiteks liiki 34 ja mõnd väiksemat lõiku 6-st. Praeguseks peaks väljaanne olema lõpetatud, kuid selle varasemad osad on ammu aegunud. Lühitabelid ilmusid viimati 1961. aastal, kauaoodatud keskmised tabelid (BS 1000M) 1985. aastal (aineloend 1988). 1993. aasta lõpul jõudis müügile BS 1000M uus versioon, nii tabelid kui ka aineloend. Inglisekeelsed tabelid on väga laialdaselt kasutusel ka väljaspool inglise keelt kõnelevaid maid ja tihti aluseks rahvuslikele UDK editsioonidele. BS 1000M-il rajaneb esimene UDK Konsortsiumi poolt müüdav standardväljaanne *Master Reference File* (MRF). BSI on konsortsiumi asutajaliige.

Ühe 1978. aasta uurimuse järgi pidas 334-st küsitletud Suurbritannia raamatukogust 270 (81%) süstemaatilist kataloogi (ülejäanud valdavalt sõnastikkataloogi). Kõige populaarsem klassifikatsioonisüsteem oli DDC, mida kasutas peaaegu 100% ava-

likke raamatukogusid [11, lk. 33]. Teisel kohal oli UDK. 1980. aastal oli Suurbritannias (ilma Põhja-Iirimaa) 2895 raamatukogu ja infotalitust, millest 640 (22%) kasutasid UDK-d [9, lk. 97].

Saksamaa raamatukogudes on UDK-d sisuavamiseks kasutatud umbes 70 aastat, kuid kasutajate ring ei ole lai. 1977. aastal Saksamaa Liitvabariigis läbi viidud küsitluse järgi pidas vaid 5 raamatukogu oma süstemaatilist kataloogi UDK järgi, kõik nad olid ülikooliraamatukogud [8]. Üldse oli vaatluse all 50 teadusraamatukogu, milles oli kokku käibel 27 erinevat liigitussüsteemi. 23 süsteemi tarvitati ainult ühes kogus — seal, kus nad välja töötatud olid. Rahvusvahelistest liigitustest oli esindatud ainult UDK. Sellise killustatuse paneb viidatud ülevaate autor Saksamaa ajalooliste traditsioonide, raamatukogundusliku individualismi ja liidumaade anakronistlike autonoomiapüüdluste süüks.

Saksa Demokraatliku Vabariigi raamatukogude tegelesid sel ajal nõukogude raamatukogu- ja bibliograafialiigituse (RBK) adapteerimisega [8, lk. 39; 15].

Mõni aasta hiljem tehtud küsitluse eesmärgiks oli uurida just UDK kasutamist. See hõlmas peale Saksamaa ka Austria ning peale raamatukogude ka teised infoasutused. Kokkuvõttes nenditakse, et UDK on mingiks otstarbeks käibel 167 asutuses Saksamaal, neist 46 infoteenistust ja 50 raamatukogu. **Austrias** on UDK-d rakendavaid asutusi 16, neist 3 infoteenistust ja 1 raamatukogu. Ainevaldkondadest, kus UDK-liigitus kõige rohkem levinud, tõusis selgelt esile tehnika [3].

Seoses Saksamaaga on huvitav märkida, et UDK on ühtseks liigitussüsteemiks kõigis Goethe instituutide raamatukogudes. Neid oli 1986. aastal 66 riigis umbes 130 [33, lk. 17].

Hoolimata UDK suhteliselt piiratud kasutamisest riigis on Saksamaa pikka aega olnud FID-i (*International Federation for Information and Documentation*) üks tugipunkte UDK rahvusvahelisel haldamisel. DIN (*Deutsches Institut für Normung*) on välja andnud saksakeelseid rahvusvahelisi tabeleid ja FID-i ametlikku publikatsiooni "DK-Mitteilungen". UDK uus keskmine väljaanne ilmus saksa keeles 1978. aastal (aineloend 1985). Täielikest tabelitest oli 1990. aastaks ilmunud määrajate osa ning liigid 3, 5, 62, 63/65, 66, 67/69 ja 7/9. Siis lõpetas DIN UDK-ga tegelemise ning publitseerimine seiskus.

Taani. 1972. aastal valis Taani teadusraamatukogude nõukogu UDK soovitatavaks liigitussüsteemiks teadusraamatukogudele. Mõni aeg hiljem asetati UDK kõrvale ka DDC, mis oli ilmutanud suurt rahvusvahelist elujõudu. 1976. aastal loodi Taani UDK-komitee. Aastatel 1978–1980 küsitletud 195 teadusraa-

matukogust pooltes kasutati omaloodud liigitussüsteeme, UDK oli üldskeemidest esikohal. Küllalt palju liigitati teadusraamatukogudes ka DK5 järgi, mis on Taani oma kümnendliigitus avalikele raamatukogudele [6]. Taanikeelsed UDK tabelid on välja antud lühendatud mahus. UDK järjekindlaid kasutajaid on *Danmarks Tekniske Bibliotek* ja Roskilde Ülikooli raamatukogu.

Rootsis ilmus esimene UDK väljaanne juba 1946. aastal. 1990. aastateni tegutses Rootsi UDK-komitee, kes tegeles tabelite ja muudatuste publitseerimisega. Viimane rootsikeelne UDK pärineb aastast 1977 muudatused aastast 1990. Praegu valmistatakse Rootsi raamatukogunduskoolis ette uusi lühitabeleid, mis peaks katma enamiku UDK-d kasutavate raamatukogude vajadused. Need vähesed erialaraamatukogud, kes vajavad täpsemaid tabeleid, võivad kasutada ingliskeelset keskmist või täielikku väljaannet [2]. Selline lähenemisviis tundub arukas, sest Rootsi UDK-kasutajaskond pole suur — peamiselt ettevõtteraamatukogud. Rohkem on levinud Rootsi rahvuslik liigitussüsteem SAB ja DDC.

Soomes võeti UDK infotalletuses kasutusele 1950. aastatel. Sõjajärgne ülesehitustöö ja maa industrialiseerimine nõudis suure hulga teadusliku ja tehnoloogilise teabe hankimist välismaalt, mida osaliselt tehti UDK vahendusel [7, lk. 43]. Niisiis juurdus UDK Soomes kõigepealt tehnika- ja majanduslase teabe süstematiseerimisel, selle abil sai anda "täpseid vastuseid täpsetele küsimustele" [10, lk. 2].

UDK positsioon Soome teadusraamatukogudes tugevnes eriti pärast seda, kui Oulu Ülikooli raamatukogu oli selle kasuks otsustanud kohe oma asutamise ajal (1959) ja Jyväskylä Ülikooli raamatukogu oma kogude avamise süsteemi uuendades (1966). Need olid esimesed suured üldteaduslikud, seega ka humanitaar- ja ühiskonnateadusi hõlmavad raamatukogud, kus võeti kasutusele UDK.

1978. aastal valitsenud seisu tähtsamate teadusraamatukogude liigituspoliitikas kajastab V Kautto artikkel [14]. Selles võetakse kokku 30 raamatukogult saadud andmed. UDK-d kasutati sisupeerelduseks 17 raamatukogus, neist 10-s ainult seda. Teised üldisemad liigitused olid märksõnaliigitus (6 raamatukogu) ja NLM (*National Library of Medicine*'i) liigitus — 4. Dewey süsteemi kasutas üks raamatukogu. Enam või vähem algpäraseid liigitussüsteeme oli 7 raamatukogul. Peale selle teatasid 9 raamatukogu omaloodud liigitustest, mida kasutati kas ainult paigutusskeemina, ainult filiaalides või mingi kindla fondiosa süstematiseerimiseks.

Ühte liigitussüsteemi kasutas 15 raamatukogu, kahte erinevat 6 ja kolme või enam 8 raamatukogu. 20 raamatukogus toimus keskliigitamine, 9-s oli see töö hajutatud.

Teine samalaadne küsitlus, mis hõlmas 64 kõrgkooli-raamatukogu (neid, mis olid ühinenud või ühinemas LINNEA-võrku), tehti 1993. aastal eesmärgiga saada ainekandjate andmebaasi (kõrgkooliraamatukogude koondkataloog) sisupegeldusreeglite koostamiseks. Kokkuvõtte näitas, et 43 raamatukogu 64 vastanust kasutas UDK-d liigituses, 21 ei kasutanud. 2 raamatukogu kavatses UDK-st loobuda, 2 selle lähemal ajal rakendada [28].

Kõiki raamatukogusid hõlmavaid küsitlusi on korraldanud *Tietopalveluseura* 1982. ja 1989. aastal. 1982. aastal oli Soomes 126 UDK-d kasutatavat raamatukogu, neist 55 (44%) asusid ettevõtete, 37 (30%) kõrgkoolide, 4 (3%) muude õppeasutuste, 8 (7%) uurimisasutuste, 12 (10%) riigiasutuste ja 7 (6%) mitmesuguste ühingute juures [18]. 1989. aastaks oli lisandunud hulk kasutajaid haldusasutustes, kuhu oli loodud uusi infotalitusi. Avalikest raamatukogudest vastas 61, et nad kasutavad UDK-d infootsingus [27].

Nende küsitluste tulemuste põhjal võib tõdeda, et UDK on Soomes oluline liigitussüsteem.

Üks tähtsamaid põhjusi UDK laiemaks levikuks 1980. aastatel on olnud üldkasutatavad andmebaasid (riikliku andmepanga KDOK ja raamatukoguvõrgu LINNEA andmebaas). Nende kirjade enamik on liigitatud UDK järgi. Teisalt teravnesid sidusotsinguga seoses ka UDK rakendamise probleemid. Selgus, et see ei vii kuhugi, kui raamatukogud jätkavad liigitamist nii, nagu ise tahavad: tulemused muutuvad kasutamiskõlbmatuks. I. Hovi väidab: "Liigituseks kulub raamatukogudes väga palju tööjõudu... Oleks olnud mõistlik, kui kasvõi murdosa sellest tööst oleks kasutatud koordineerimiseks, ühiste tööviiside kokkuleppimiseks ning ainealoendite ja normatiivsete registrite koostamiseks" [10, lk. 4].

Soome keeles on olemas UDK lühitabelid (3. väljaanne 1983, parandustega lisatrükk 1991) ja ainealoend (1985). Uut ainealoendit hakati tegema 1991. aastal, kuid see töö jäi rahapuuduse tõttu pooleli.

Portugalis on UDK valdav raamatukoguliigitus. Esimene väljaanne ilmus 1954. aastal. UDK-d kasutatakse nii rahvusraamatukogus kui ka kõigis universaalsetes teadusraamatukogudes. 1987. aastast on käigus PORBASE, Portugali raamatukogude

koondkataloog, milles UDK on esmane ja kohustuslik sisupeegeldusvahend. Liigitusühtsust aitab tagada autoriteetfail.

1996. aastaks on kavandatud ilmuma uus portugalikeeelne UDK CD-ROM-il [21].

Lääne-Euroopa riikidest on UDK küllalt hästi tuntud veel Belgias, Hispaanias, Hollandis, Itaalias, Norras, Prantsusmaal. Portugali- ja hispaaniakeelsed tabelid on mõjutanud UDK-d rakendama ka Ladina-Ameerika riike. Kaheksakümnendatel aastatel sai valmis itaaliakeeelne täielik väljaanne. Kõige uuematest publikatsioonidest Lääne-Euroopa keeltes on nimetamata veel Belgias ilmunud prantsuskeelsed keskmine (1990–1993) ja lühiväljaanne (1994) ning hispaaniakeeelne lühiväljaanne (1991).

Ida-Euroopa riikides on UDK levikut muu hulgas soodustanud sotsialismile omane riiklik ja riikidevaheline tsentraliseerimine NSV Liit püüdis sotsialismileeri ühtse infootsikeelena juurutada RBK-d, kuid eriti Poola, Ungari, Rumeenia ja Serbia olid selle vastu [15, lk. 21]. Nendes maades olid UDK positsioonid tugevad juba seitsmekümnendatel aastatel.

Ungari raamatukogud ja infoasutused on kõik UDK-le orienteerunud. Täielikud tabelid ilmusid aastatel 1968–1972 (42 vihikut, enamikku neist on uuendatud) [37]. Uusim lühiväljaanne ilmus 1991–1992.

Tšehhis ja Slovakkias on UDK samuti kaua aega tuntud ja kasutusel. Erialatabeleid ilmus juba kahekümnendatel ja kolmekümnendatel aastatel, esimene terviklik UDK lühiväljaanne 1948. aastal. Täielikest tabelitest, mida avaldati nii tšehhi kui ka slovaki keeles, ilmus aastatel 1968–1990 35 vihikut. Järjekindlalt lisandus neile täiendusi ja parandusi. Aastatel 1994–1995 publitseeriti tšehhikeeelne lühendatud UDK, üks esimesi UDK Konsortsiumi litsentsiga väljaandeid. Praha Rahvusraamatukogus on käimas kogu MRF-i tõlkimine, tšehhikeeelne versioon pidi valmima 1996. aastal [17].

Peale raamatukogude ja rahvusbibliograafia on UDK näidanud oma elujõudu ka mõnes kaheksakümnendatel aastatel Slovakkias loodud erialases infootsisüsteemis. Liigitust on neis kasutatud suurte infomassiivide esialgseks liigendamiseks, kusjuures spetsiifilisemad päringud saab teha deskriptorite või võtmesõnadega [29, lk. 133–134].

Poola raamatukogudes on UDK enim kasutatav liigitussüsteem. Ka infotalitustes leidis ta laia rakendust — 1984. aastal 166 asutusest vähemalt 61-s [41]. 1980. aastate alguses töötati välja oma universaalne klassifikatsioonisüsteem, kuid see pole riigis valdavaks saanud. Poola keeles on ilmunud igat tüüpi UDK

tabeleid, neist uusim on aastatel 1993–1994 publitseeritud lühiväljaanne.

Leedus võeti neljakümne aastal küll suund DDC ja UDK kasutamisele, kuid peagi tuli hakata järgima NSV Liidus kehtestatud nõudeid. Osades raamatukogudes elustus UDK 1960. aastatel venekeelsete tabelite järgi, suurem osa aga läks üle RBK-le. 1988. aastal otsustati välja anda leedukeelne UDK. Aluseks võeti Poolas ja teistes sotsialistlikes riikides ilmunud tabelid. 1991. aastal sai valmis lühiväljaanne rahvaraamatukogudele, rohkemaks esialgu jõudu ei jätkunud [25]. Praegu on Leedu Rahvaraamatukogus valinimas MRF-i järgi tõlgitav UDK keskmine väljaanne.

Kirjutistest nopitud andmete, kuuldud ettekannete ja isiklike vestluste põhjal võib öelda, et UDK on kujunenud või kujunemas ühtseks raamatukoguliigituseks kõigis Ida-Euroopa riikides. Seda tõendab kasvõi fakt, et 1995. aasta sügiseni UDK Konsortsiumi poolt müüdnud 20 MRF-i litsentsist 11 on omandanud endised sotsialismimaad [20]. Infoteenistuste kohta sedasama väita ei saa. Mitmes kümnend aastat tagasi kirjutatud artiklis tõdetakse UDK taandumist infoasutustest seoses infotöö automatiseerimisega [36; 41]. Praeguseks on paljud Ida-Euroopa infoteenistused üldse tegevuse lõpetanud, sest nende riiklik ülalpidamine on lakanud ning tarbijad eelistavad läänest saadavat erialateavet.

Venemaa (NSVL). Vene raamatukoguhoidjad hakkasid kümnendliigituse vastu huvi tundma juba möödunud sajandi lõpus, pärast 1895. aastal Brüsselis toimunud rahvusvahelist bibliograafikonverentsi. DDC ja "Brüsseli liigitus" millest kasvaski välja UDK, nende algupärasel kujul laia rakendust küll ei leidnud, kuid koostati mitmeid lihtsamaid detsimaalsüsteemil põhinevaid liigitusi, näiteks P. Bogdanovi poolt [40, lk. 389–398].

Nõukogude valitsus oli huvitatud ühtse liigituse kehtestamisest kõigis raamatukogudes, mis võimaldaks personali tsentraliseeritud väljaõpet, keskliigitamist, raamatukogude varustamist trükikaartidega ning muidugi ideoloogilist suunamist ja kontrolli. 1921. aastal võeti vastu N. Krupskaja allkirjaga määrus rahvusvahelise kümnendliigituse rakendamisest kohustusliku süsteemina kõikides VNFSV raamatukogudes. See määrus jäi tegelikult ellu viimata. Avalike ja teadusraamatukogude liigitussüsteemid arenesid algusest peale lahus. Rahvaraamatukogude tabeleid redigeeriti põhjalikult, arvestades nõukogude ühiskonna reaale. Säilitades osaliselt kümnendsüsteemi vormi, oli nende sisu samal ajal hoopis teistsugune. Teine variant — teadusraamatukogudele — arenes eraldi, kuid ei võtnud arvesse ka FID-i tegevust ning

seisukohti [30, lk. 69]. NSV Liidu Ministrite Nõukogu määrusega sai UDK 1962. aastal kohustuslikuks teaduslikele ja tehnikaraamatukogudele ning infotalitustele. Pärast seda algas ka koostöö FID-iga. Siiski tarvitatakse UDK-d Venemaal enamasti ainult loodusteadusliku ja tehnilise kirjanduse liigitamiseks.

1988. aasta andmetel kasutati UDK-d NSV Liidus rohkem kui 50 000 asutuses, sealhulgas 14 600 infotalitust ja 11 200 teadus- ja tehnikaraamatukogu. Küllalt hästi toimis keskliigitamine ja trükikaardisüsteem [36, lk. 13]. UDK Konsortsiumi loomise alg-etapil olid liikmeks saamisest huvitatud ka NSV Liidu esindajad [38], kuid on hiljem siiski kõrvale jäänud.

Venekeelsete tabelite viimane, 3. väljaanne (1979–1986) esitab kõige põhjalikumalt pealiigid 5 ja 6. Kõige rohkem lühendusi ja moonutusi on pealiikides 1 ja 2, kuhu nurksulgudes on lisatud formuleeringuid, mille abil saaks liigitada marksistlik-leninlikku kirjandust. Seda võtet kritiseerib teravalt E. Sukiasjan, märkides, et need täiendused lahknevad UDK etalonist ega rahulda samal ajal ka nõukogude raamatukogude vajadusi [39, lk. 60].

Arengumaades kasutatakse küllalt palju Euroopas ja Ameerikas loodud klassifikatsioonisüsteeme, eriti DDC-d ja UDK-d. Kuid need on Lääne kultuuri kesksed ega arvesta vajalikul määral Ida tsivilisatsioonide omapära. Eriti on see maksev humanitaar- ja sotsiaalteaduste suhtes, nagu näiteks religioon, keeled, kirjandus, ajalugu ja geograafia. Arengumaade raamatukogutegelased on kritiseerinud universaalklassifikatsioone, sealhulgas UDK-d, Aasia ja Aafrika ainestiku puuduliku, vigase, aegunud ja tagaplaanile surutud esituse pärast [31]. Samal ajal püüavad nad ise neid süsteeme täiendada vajalike liikidega. Pakistanis näiteks on laiendatud UDK-d nii, et sellega saab rahuldavalt liigitada islamiuurimusi. Vastavad tabelid hõlmavad 84 lehekülge [4, lk. 264]. Rahvuslike liigitussüsteemide väljatöötamine on küllalt kallis ning selleks ei jätku arengumaadel raha ega asjatundjaid.

Liigitusprobleemidest **Aafrika** raamatukogudes oli ettekanne IFLA Moskva konverentsil 1991. aastal [5]. Selles märgitakse, et UDK-d kasutatakse põhiliselt Aafrika prantsuskeelsete osade raamatukogudes, ingliskeelsetes raamatukogudes on käibel DDC ja LCC (*Library of Congress Classification*). Siiski leidub ka näiteks ingliskeelses Nigeerias 16 raamatukogu, kus rakendatakse UDK-d [35].

Jaapani kõige populaarsem liigitussüsteem on Jaapani kümnnendliigitus, mille loomisel on kasutatud C. A. Cutteri ideid.

UDK-d kasutas 1986. aastal 264 erialaraamatukogu (14% raamatukogude üldarvust) [12]. Maa suurim teadus- ja tehnika-informatsiooni sisaldav elektronandmebaas JICST samuti kõik rahvuslikud standardid on liigitatud UDK järgi [24, lk. 231].

Jaapanikeelset UDK keskmist väljaannet (1984) peetakse üheks õnnestunumaks. Jaapan on esindatud ka UDK Konsortiumis. Mõned jaapani uurijad on tõsiselt tegelnud UDK automatiseerimise probleemidega.

Teises Aasia suurriigis **Hiinas** on UDK küll tuntud, kuid vähe praktikas kasutatud. Olulisemad on rahvuslikud liigitussüsteemid, mille arendamisel on teatud määral eeskuju võetud UDK sünteesilisusest, kuid enam toetutakse koolonliigitusele (CC) [34].

Koolonliigituse sünnimaal **Indias** on CC levikult alles kolmandal kohal. Enamik raamatukogusid kasutab DDC-d (65%) ja UDK-d, kuigi koolonliigitus võimaldab indoloogiasse puutuvat ainet liigitada üksikasjalikumalt [26].

Universaalse kümnendliigituse levikut kokku võttes tuleb märkida, et UDK-l on 100 000 institutsionaalset ja individuaalset kasutajat umbes 60 riigis. Paljudes maades, eriti Ida-Euroopas, on ta ühtseks liigitussüsteemiks kõigis või enamikus raamatukogudes. Tabelid on täielikus, keskmises või lühendatud mahus olemas vähemalt 25 keeles. Viimastel aastatel on ilmunud palju uusväljaandeid, mis annab tunnistust UDK jätkuvast populaarsusest. UDK-keelte hulka kuuluvad inglise, saksa, prantsuse, hispaania, portugali, itaalia, vene, rootsi, taani, soome, hollandi, poola, leedu, ungari, tšehhi, slovaki, serbohorvaadi, sloveeni, makedoonia, bulgaaria, rumeenia, jaapani, hiina, heebrea, araabia ja eesti keel. Peaks olema piisavalt esinduslik seltskond, et sellega peatselt liituda.

Kirjandus

1. Aasmets, M. Saja-aastane UDK sajandilõpu Eestis // Raamatukogu. 1995. Nr. 4. Lk. 14–16.
2. Benito, M. The UDC in Sweden // Extensions and Corrections to the UDC. 1995. Vol. 17. P. 83–84.
3. Burkart-Sabsoub, M., Wersig, G. Nutzung der DK in der Bundesrepublik Deutschland und Österreich // DK-Mitteilungen. 1982. Jg. 26. Nr. 3. S. 9–12.
4. Cataloging and classification of non-Western material: concerns, issues and practices. Phoenix, 1980. 368 p.

5. Diongue-Diop, M. Les problèmes de classification et d'indexation dans les bibliothèques Africaines // International Cataloguing and Bibliographic Control. 1992. Vol. 21. N. 3. P. 44–47.
6. Friis-Hansen, J. B. Subject retrieval in Denmark, 1975–1985: a survey // International Classification. 1985. Vol. 12. N. 2. P. 66–71.
7. Haarala, A.-R. The role of UDC in Finnish classification policy // International Cataloguing and Bibliographic Control. 1991. Vol. 20. N. 3. P. 43–46.
8. Heinrich, G. Klassifikatorische Sacherschliessung in deutschen Bibliotheken // Kooperation in der Klassifikation 2. Frankfurt/[Main]. 1978. S. 33–53. (Studien zur Klassifikation; 3).
9. Hindson, R. UDC in the UK: a report on the 1979/1980 survey // Aslib Proceedings. 1981. Vol. 33. N. 3. P. 93–101.
10. Hovi, I. Tieteellisen kirjallisuuden luokitus // Signum. 1987. V. 20. No. 1. S. 1–5.
11. Hunter, E. J. The United Kingdom contribution to subject cataloguing and classification since 1945 // International Cataloguing. 1987. Vol. 16. N. 3. P. 31–34.
12. Ishiyama, H. The trend of classification in Japan // International Cataloguing. 1986. Vol. 15. N. 3. P. 31–33.
13. Janke, E. Die Dezimalklassifikation in der Normeninformation // DK-Mitteilungen. 1985. Jg. 29. Nr. 6. S. 19–24.
14. Kautto, V. Suomen suurimpien tutkimuskirjastojen luokitus- ja indeksijärjestelmät ja atk-systeemin tiedonhakuun liittyvät toivomukset // Signum. 1978. V. 11. No. 7. S. 142–146.
15. Kehman, A. Nõukogude raamatukogu- ja bibliograafiklassifikatsioon levik sotsialismimaades (Saksa DV raamatukogude kogemused): diplomitöö. Tallinn, 1977 — Käsikiri TPÜ-s.
16. Kirsel, A.-M. UDK eestikeelse väljaande saamisloost // Eesti Raamatukoguhoidjate Ühingu aastaraamat 1995. Tallinn, 1996. Kd. 7. Lk. 68–75.
17. Kofnovec, L. The UDC in the Czech Republic // Extensions and Corrections to the UDC. 1995. Vol. 17. P. 77–78.
18. Kuuteri, R. UDK-luokitusjärjestelmää käyttävät kirjastot Suomessa // Signum. 1982. V. 15. No. 10. S. 232.
19. Library bibliographic networks in Europa: a LIBER directory. 2nd ed. The Hague, 1992.
20. Licensees to the UDC Master Reference File // Extensions and Corrections to the UDC. 1995. Vol. 17. P. 90.
21. Lopes, M. I. The UDC in Portugal: background and prospects // Extensions and Corrections to the UDC. 1995. Vol. 17. P. 81–82.
22. McIlwaine, I. C. Guide to the use of UDC: an introductory guide to the use and application of the Universal Decimal Classification. The Hague, 1993.

23. McIlwaine, I. C. Present role and future policy for UDC as a standard for subject control // Standards for the international exchange of bibliographic information. London, 1991. P. 151–156.
24. McIlwaine, I. C. Seminar on "UDC and its Development Towards the 21st Century" // Knowledge Organization. 1994. Vol. 21. N. 4. P. 231–232.
25. Procopciuk, M. Classification and indexing policy in Lithuania: [loeng Oulu Ülikooli raamatukogunduse ja infoteaduse osakonnas 18. 01. 1993]. — Konspekt autori valduses.
26. Raju, A. A. N. Indology — problems in its classification and organization in libraries with special reference to Universal Decimal Classification (IME, 1985) and Colon Classification (7th edition, 1987) // Tools for knowledge organization and the human interface 2. Frankfurt / Main, 1991. P. 124–133. (Advances in knowledge organization; 2).
27. Seppänen, M.-L. UDK-luokituksen käytön kartoitus 1989 // Tietopalvelu. 1989. V. 4. No. 8. S. 21.
28. Sisällönkuvailukyselyn alustava yhteenveto. 1993. — Käsikiri, paljundus autori valduses.
29. Staněiková, P. Present state and future development of information retrieval languages in Czechoslovakia // International Classification. 1984. Vol. 11. N. 3. P. 133–138.
30. Sukiasyan, E. R. Classification practice in the USSR: current status and development trends // International Classification. 1988. Vol. 15. N. 2. P. 69–72.
31. Swaydan, N. M. The universal classification and the needs of libraries in developing countries // Universal classification 1: subject analysis and ordering systems. Frankfurt/[Main], 1982. P. 321–328. (Studien zur Klassifikation; 11).
32. Traiser, W. Anwendung der universellen Dezimalklassifikation in Nationalbibliographien // Wissenorganisation im Wandel: Dezimalklassifikation — Thesaurusfragen — Warenklassifikation. Frankfurt / Main, 1988. S. 27–42. (Studien zur Klassifikation; 18).
33. Wimmer, F. Die universelle Dezimalklassifikation (UDC) als Instrument einer gemeinsamen Sacherschließung in einem Bibliotheksverbund // DK-Mitteilungen. 1986. Jg. 30. Nr. 5. S. 15–18.
34. Xiao, Yan. Modern development of classification: research and practice in the People's Republic of China // International Classification. 1992. Vol. 19. N. 1. P. 10–14.
35. Гривз М. А. Универсальная десятичная классификация и ее использование в Нигерии // Международный форум по информации и документации. 1987 Т. 12. № 3. С. 25–30.
36. Кристальный Б. В. и др. УДК в государственной системе научно-технической информации // Научно-техническая информация, Сер. 1. 1988. № 11. С. 13–15.

37. Мартъян Д., Барта Г. Деятельность венгерского Редакционного совета по УДК в 1982–1985 гг: состояние венгерского издания УДК // Международный форум по информации и документации. 1986. Т. 11. № 4. С. 14–18.
38. Сейфуль-Мулюков Р. Б. Новая политика Международной Федерации по документации в отношении УДК // Научно-техническая информация, Сер. 1. 1991. № 10. С. 22–23.
39. Систематический каталог: практическое пособие / Сост. Э. Р. Сукиасян. Москва, 1990.
40. Шамурин Е. И. Очерки по истории библиотечно-библиографической классификации 2. Москва, 1959.
41. Ябженская Э., Сыцибор Э. Обзор языков индексирования, используемых в научно-информационных органах ПНР // Международный форум по информации и документации. 1987. Т. 12. № 2. С. 12–13.

THE WORLDWIDE USE OF THE UNIVERSAL DECIMAL CLASSIFICATION

Sirje Nilbe

Summary

The most accepted classification system in the Estonian research libraries has been the Universal Decimal Classification (UDC). In the autumn of 1995 the centennial of the UDC was celebrated. In the last years the management of the UDC has changed and the classification is now available in a machine-readable format. This database called MRF favours the future development and use of the UDC.

The UDC is extremely versatile by nature. This makes it appropriate for a wide variety of situations. The UDC can serve as:

- a shelf arrangement scheme in the libraries;
- a basis of the classified card catalogues;
- a tool for subject searching in online catalogues and other databases;
- a framework for subject arrangement in indexing services publications covering journal articles, research reports and similar materials;
- a subject indicator in national bibliographies both in printed and digital form;
- a tool for thematic arrangement of the networked information resources.

Mostly, the available about the use of the UDC data refer to European countries.

Great Britain has the leading position in publishing and distributing the UDC tables in English (British Standards Institution being the founding member of the UDC Consortium). 22% of libraries and information agencies in Great Britain were using UDC in 1980.

The Deutsches Institut für Normung managed the UDC tables in German till 1990. There were 167 institutions using UDC in Germany in 1982.

In Denmark and Sweden the abridged versions of the UDC are being used in research and special libraries, in public libraries the national classification systems are more popular.

In Finnish research libraries the UDC has a strong position since the sixties. The majority of records in various bibliographic databases have UDC numbers. The UDC abridged tables in Finnish have been published several times.

In Portugal the UDC is the dominating library classification. In the union catalogue PORBASE every record contains a UDC number.

In Central and Eastern Europe the classification policy and practice has been more centralized than in Western Europe. UDC has had the dominating position in the libraries of Hungary, Czech Republic, Slovakia, Poland, etc. There is a great variety of UDC tables in those languages. Remarkably, the postsocialist countries have obtained 11 of the first 20 licences to the UDC MRF for publishing the new editions.

In Russia the new classification policy has not been formed yet. Since 1962 the UDC has been mandatory for scientific and technical libraries and information agencies in the Soviet Union. The tables in Russian were used in 50 000 institutions. Representatives of the Soviet Union had strong positions in the international UDC bodies. Since the transition of the management of the system to the UDC Consortium, the Russian specialists do not take part in the revision work.

Libraries of the Asian and African countries are widely using the classification systems created in Western countries, especially DDC and UDC. There are many local entities not represented in these systems but the development of a new universal classification is very costly. Only a wealthy country like Japan or a very big country like China have been able to develop its own national classifications.

UDC has 100 000 institutional and individual users in 60 countries. In many countries, especially in Central and Eastern Europe, the UDC serves as a shared subject indexing tool. The tables at abridged, medium or full level are published in 25 languages. Several new editions have recently come out. Very likely, the first Estonian edition will be in use in our libraries quite soon.

**Estonian Librarians Association
Общество библиотекарей Эстонии**

**National Library of Estonia
Национальная библиотека Эстонии**

INDEPENDENCE AND LIBRARIES НЕЗАВИСИМОСТЬ И БИБЛИОТЕКИ

**Papers
of the 5th Congress of Baltic Librarians
October 21–22, 1996
Tallinn, Estonia**

**Доклады
5го Конгресса библиотекарей Балтийских стран
21–22 октября 1996
Таллинн, Эстония**

Tallinn 1996 Таллинн

CONTENTS / СОДЕРЖАНИЕ

PLENARY SESSION / ПЛЕНАРНОЕ ЗАСЕДАНИЕ

Tiiu Valm The Development of Library Activities in Estonia in 1990-1995 / Развитие библиотечного дела Эстонии в 1990 – 1995 гг	7
Audronė Glosienė Librarianship in Lithuania in 1990-1995	13

PARALLEL SESSIONS / ПАРАЛЛЕЛЬНЫЕ ЗАСЕДАНИЯ

Research and special libraries / Научные и специальные библиотеки

Maret-Ene Klaus О координации комплектования между крупнейшими библиотеками Эстонии / Acquisitions Co-ordination in Estonian Research Libraries	21
Mihkel Volt On Acquisition of the Baltica Publications in Estonian Research Libraries Today / О комплектовании фондов с балтикой в эстонских научных библиотеках сегодня	24
Rima Mockapetrienė Координация комплектования зарубежной периодики в Литве	27
Viesturs Zanders “ <i>Baltische Zentrale Bibliothek</i> ” как источник изучения культурных связей балтийских народов	30

Public libraries / Массовые библиотеки

Elviine Uverskaja Предпосылки и условия развития публичных библиотек Эстонии / Estonian Public Libraries – Assumptions and Opportunities of Change	32
Dzidra Šmita Public Libraries in Latvia Today	39
Rutt Enok Библиотека — центр информации и культуры / Public Library as centre of Culture and Information	42
Aldona Kerpytė Публичная библиотека — центр учебной информации и культурной деятельности	45
Vida Garunkštytė Публичная библиотека: ее место в сегодняшней жизни	48

Children and school libraries / Детские и школьные библиотеки

Anne Rande Каждый читатель начинается с детства / All Readers are Born as Children	50
Virginija Liutikaitė Как изменилась детская библиотека в годы независимости	55
Silvija Tretjakova Children and School Library Situation in Latvia	58

Mall Saul Estonian School Libraries Yesterday, Today and Tomorrow / Библиотеки Эстонских школ в настоящем и будущем	61
Dženija Dzirkale Школьные библиотеки Латвии: настоящее положение и тенденции развития	66
Birutė Vaitiekūnienė Школьная библиотека в Литве и ее проблемы	71

Professional education and development / Профессиональное образование

Elena Mačevičiūtė Library and Information Science at the Vilnius University	72
Baiba Sporāne Library Education in Latvia: Present and Future	77
Povilas Abarius Становление новой системы непрерывного обучения библиотечных работников Литвы	82
Sirje Virkus Distance Education in Library and Information Science Education in Estonia / Дистанционное обучение в системе библио- течного и информационного образования Эстонии	84

Standardisation / Стандардизация

Marje Aasmets, Arda-Maria Kirsal The UDC MRF in Estonian / Машиночитаемая версия УДК (Мастер Референс Файл) на эстонском языке	92
Regina Varnienė Стандарты и машиночитаемые форматы LIBIS / Интегрированная информационная система библиотек Литвы	97
Silvi Metsar ISSN and ISBN in Estonia / ISSN и ISBN в Эстонии	101
Ljubov Buckienė Библиографические стандарты, их перевод на литовский язык, практическое применение	105
Sirje Nilbe Standardisation in Subject Indexing – Possibilities and Activities / Стандартизация предметизации — возможности и их реализация	107

Co-ordination and co-operation / Координация и кооперация

Birutė Railienė International Relations of the Lithuanian Librarians' Association	112
Anne Valmas The Aim of Estonian Librarians Association to Promote the Co- operation Between Libraries / Роль Общества библиотекарей Эстонии в координации межбиблиотечной деятельности	115
Aija Janbicka The Association of Latvian Academic Libraries and its Activities	119
Emilija Banionytė TEMPUS Project in Lithuanian Academic Libraries	123
Ülle Must Information Seeking in Electronic Environments: Reflection of Baltic Development in Electronic Databases / Информационный поиск в электронной среде: рефлексия развития балтийских библиотек в электронных базах данных	129

STANDARDIZATION IN SUBJECT INDEXING: POSSIBILITIES AND ACTIVITIES

Sirje Nilbe
Tartu University Library

The cooperation and automation projects in Estonian libraries have stimulated the looking for possibilities to harmonize subject analysis in our catalogues and other databases. The indexing tools as the UDC Estonian Edition and the Estonian Universal Thesaurus are being developed. But the shared indexing languages alone do not guarantee the necessary standardization level. In this paper the possibilities and activities for harmonizing subject analysis, especially verbal subject indexing, are overviewed with the help of the international guidelines, produced by ISO and IFLA.

Indexing process

The indexing process has four stages, although these tend to overlap in practice:

1. Examining the document and establishing its subject content.
2. Identifying the principal concepts in the document.
3. Expressing these concepts in the terms of the indexing language.
4. Assignment of index terms to the document.

The first two of these stages taken together describe the process of conceptual or intellectual analysis, and the third is often known as the translation stage.

Conceptual analysis, first and foremost, involves deciding what a document is about — that is, what it covers. Most of the writings on content analysis are theoretical (2; 4) or experimental (1; 5). Various writers pointed to humans' limited ability to objectively and consistently determine aboutness. In Williamson's (12) opinion "it would be difficult and virtually impossible to achieve standardization of the process of content analysis" However, ISO 5963-1985 *Documentation — Methods for examining documents, determining their subjects, and selecting indexing terms* (7) provides some useful guidelines for this aspect of the indexing.

By examining the document a complete reading is often impracticable. The standard gives a checklist of the important parts of the text that need to be considered carefully: title; abstract, if provided; contents list; introduction, opening phrases of chapters and paragraphs, and the conclusion; etc. Warnings are given against indexing from title alone (which might be insufficient and perhaps misleading) or from an abstract (which may be inadequate). It is mentioned that non-print documents call for different procedures. These guidelines might seem trivial but I think they are not — how often they are ignored!

After examining the document, the indexer should follow a systematic approach to the identification of those concepts which are essential elements in a description of its subject. The choice of those concepts depends on the purpose for which the indexing terms will be used. The identification of concepts may also be affected by the item being indexed. The main criterion should always be the potential value of a concept as an element in the expression of the subject of the document and in its retrieval.

These statements in the standard indicate to fact that exhaustivity and specificity of the identification of concepts inevitably varies from one database to another. There is no one "correct" set of index terms for any item. It means that the more detailed rules for document

analysis are not a matter for an international or national standard. They should be established at the database level.

For the correct translation the identified concepts into indexing terms the indexer should be familiar with the indexing language. In particular, he or she should be aware that it may impose certain constraints. For example, a prescribed list of subject headings may not permit the exact representation of a concept encountered in a document. In this case, the concept should be represented temporarily by more general term, the new concept being proposed as candidate for later addition. In general, the most specific term available in the indexing language should be selected to represent a given concept.

Indexing languages

The quality and consistency of indexing depend not only on the qualification and expertise of the indexer but also on the quality of the indexing language.

The verbal indexing languages are in general of two kinds — subject heading lists and thesauri. Particularly in English and American this distinction has been conventional. Subject headings originate from library card catalogues, they are designed for pre-coordinated indexing of book and periodical collections often of universal scope. Thesauri have arisen from indexing and abstracting agencies at dawn of the electronic bibliographic databases, they are designed for post-coordinated indexing mostly for scientific articles, reports and dissertations of restricted subject domain. In truth, the distinction between the two types has recently become somewhat confused because many subject heading systems have included some structural features characterizing thesauri, for example references BT, NT, RT.

Thesauri have a relatively long history in the development of standards (9). As early as in 1967 Committee on Scientific and Technical Information in USA set up the first guidelines for thesaurus construction. In the seventies, the international bodies Unesco and ISO published their first guidelines. Today the main international authorities in this field are ISO 2788–1986 *Documentation — Guidelines for the establishment and development of monolingual thesauri* (6) and ISO 5964–1985 *Documentation — Guidelines for the establishment and development of multilingual thesauri* (8). Remarkable set of national standards for thesaurus construction is developed on this basis, for example in Great Britain, Germany, France, USA, Finland, etc.

Standards provide rules for the four basic components of thesauri — vocabulary, structure, display and management. The problems relating to vocabulary include choice of terms, scope notes and definitions, decisions to use singular or plural, what classes of words may serve as descriptors, to what extent to allow compound (precoordinate) terms, how to deal with synonymy, polysemy and homography. The rules for the structure guide the organization of terms by semantical relationships (the equivalence, hierarchical and associative relationships as the main of them). Thesaurus display must make terms and their relationships explicit and easy accessible. Guidelines for the management include usually a short methodology of compilation, editorial matters, updating, etc.

As mentioned above, thesauri have served mostly for indexing scientific documents in agencies producing online databases. But there is no reason why they should not also be used in online catalogues and national bibliographies. It is possible to apply the basic principles of ISO 2788 to any indexing language. We have some experiences in this field in Estonia. In the National Library the Estonian Universal Thesaurus in printed format is being developed according to ISO 2788. At Tartu University Library the online thesaurus for indexing and searching of our online catalogue is being composed from free indexing terms. From the standardization point of view, the parallel developing of two separate thesauri is undoubtedly disadvantageous but this is the life — the Universal Thesaurus was only at its initial stage when we began indexing for our online catalogue in November 1994.

For constructing subject headings there is no international and, insofar as I succeeded in determining, national standards. Standardization of controlled vocabularies in online

catalogues is usually guaranteed by common use of a standard subject heading list, accompanied with some kind of application manual. It is typical that these tools are compiled and maintained in national libraries who are responsible for indexing national printed word. Many lists become "standards" in their own right through cooperative use and through their application in machine-readable records which are widely distributed to other libraries. The best example is *Library of Congress Subject Headings* what is widely used outside the United States, either in original form or modified in some way.

A crucial factor in connexion with the national universal subject heading languages is heterogeneous target audience. The level of a language may be judged by public libraries as too learned and by university and specialized libraries as too general. Cross-references between "learned" and "ordinary" terms can help here.

The main international body concerned with standardizing of subject analysis in library catalogues has been IFLA Section on Classification and Indexing. In the field of verbal indexing languages the Section has produced *Guidelines for Subject Authority and Reference Entries* (3).

These guidelines set out rules for establishing terminology and structuring relationships among terms. They are meant to be used for print as well as for online authority and reference entries and were developed to facilitate the international exchange of authority entries. The data elements to be included in an authority entry are defined, and an overall structure for the entry is provided. A standard order for the elements within the entry is established.

The *Guidelines* deal with the structure of the entries; they do not prescribe the actual form of headings, references, or notes. Decisions on those matters are left to the national bibliographic agencies responsible for authority entries.

We have no practical experience with these guidelines yet because we have no common authority files at the national level.

From 1990 the IFLA Section on Classification and Indexing's Working Group have worked at the document *Principles Underlying Subject Heading Languages*. In 1995 the draft version (11) was prepared. Unfortunately, the distributing of the draft document has been limited. It is not available in Estonia.

The aims of the *Principles* are in words of Lopes (10) to facilitate subject access to information on an international level; to assist in developing subject heading languages (SHLs) by stating what is meant by a good SHL and what desirable construction and application principles are for such languages; to promote understanding of different SHLs by identifying commonalities underlying them and providing a structure for their comparative study; to provide a theoretical rationale for particular standards or guidelines for SHL construction and application. These purposes unfold the *Principles* tend to be the first unified conceptual framework for the whole process of subject indexing.

Part I in this document provides background, definitions and principles, part II includes a survey of each principle, with illustrating statements and examples taken from Canada, France, Germany, Iran, Norway, Poland, Portugal, Spain and USA. I think it is worth to wish the speedy finishing of this document.

While it should be possible to reach some international understanding about conceptual foundations of subject alphabetical retrieval systems, the terminological and language barriers are essential in international information exchange. The lack of multilingual facilities has been partially alleviated by the international use of certain languages. Nevertheless, while this may be true for science and technology, it is hardly so for the social sciences and humanities, where different languages are by themselves, in words of Lopes (10), "repositories of specific knowledge pertaining to local and regional culture" To cross the language barriers the tools as multilingual thesauri, classification systems and switching languages should be a matter for international standardization in the near future.

References

1. Chu, C. M., O'Brien, A. Subject analysis: the critical first stage in indexing // Journal of Information Science. 1993. Vol. 19 No. 6. P 439 – 454.
2. Farrow, J. Indexing as a cognitive process // Encyclopedia of library and information science. Vol. 53 (Suppl. 16). New York, 1994. P 155 – 171.
3. Guidelines for subject authority and reference entries. München, 1993. (UBCIM Publications – New Series; Vol. 12).
4. Hutchins, W. J. The concept of 'aboutness' in subject indexing // Aslib Proceedings. 1978. Vol. 30. No. 5. P 172 – 181.
5. Iivonen, M. Indeksintituloksen riippuvuus indeksintympäristöstä. Tampere, 1989. (Tampereen yliopiston kirjastotieteen ja informatiikan laitoksen tutkimuksia; 26).
6. ISO 2788: Documentation — Guidelines for the establishment and development of monolingual thesauri. 2nd ed. 1986.
7. ISO 5963: Documentation — Methods for examining documents, determining their subjects, and selecting indexing terms. 1985.
8. ISO 5964: Documentation — Guidelines for the establishment and development of multilingual thesauri. 1985.
9. Krooks, D. A., Lancaster, F. W. The evolution of guidelines for thesaurus construction // Libri. 1993 Vol. 43. No.4. P 326 – 342.
10. Lopes, M. I. Principles underlying subject heading languages: an international approach // 61st IFLA General Conference 20 – 26 August 1995 Istanbul. Booklet 4. P 48 – 53
11. Principles underlying subject heading languages (SHL's): (draft document) / IFLA Section on Classification and Indexing, Working Group on Principles Underlying Subject Heading Languages. May, 1995
12. Williamson, N. J. Standards and standardization in subject analysis systems: current status and future directions // Subject indexing: principles and practices in the 90's. München, 1995. P 278 – 291. (UBCIM Publications – New Series; Vol. 15).

СТАНДАРТИЗАЦИЯ ПРЕДМЕТИЗАЦИИ: ВОЗМОЖНОСТИ И ИХ РЕАЛИЗАЦИЯ

Сирье Нильбе

Библиотека Тартуского университета

В докладе рассматриваются возможности применения стандартизации при предметизации как на уровне отдельных баз данных, так и на национальном и международном уровне и дается обзор опубликованных по этой теме международных стандартов и других руководящих документов.

Процесс предметизации регламентируется стандартом МОС (7). Так как анализ содержания документов в значительной степени представляет собой интеллектуальную деятельность субъективного характера, стандартизация его является почти невозможной. Вышеупомянутый стандарт дает только общие руководящие указания по определению содержания документов и их предметизации. Более детальные правила должны разрабатывать сами библиотеки и другие документационные центры в зависимости от характера обрабатываемого материала и потребностей пользователей.

В большем количестве опубликованы руководства, касающиеся создания информационно – поисковых языков. Составление тезаурусов регламентируют стандарты МОС (6; 8). Они касаются основных компонентов тезаурусов – словник, структура, дисплей и ведение. В Национальной библиотеке Эстонии составляется по ISO 2788 универсальный тезаурус для унификации предметизации во всех библиотеках Эстонии. Онлайн – тезаурус интегрированный в электронный каталог Библиотеки Тартуского университета построен по тому же стандарту.

Для создания список предметных рубрик специальных стандартов не издавалось, но сами они становились стандартами, в случае, если они получали достаточно широкое распространение. В последние годы активизировалась разработка общенациональных списков предметных рубрик в связи с созданием электронных сводных каталогов. Теперь проблемой является необходимость унификации предметизации на международном уровне, поскольку инфообмен перешел национальные границы. Под эгидой ИФЛА было издано руководство по международному обмену авторизованных предметных рубрик (3).

Задачей стандартизации предметизации на международном уровне в ближайшее время является поиск возможностей преодоления языкового барьера между различными библиографическими системами.



1/1995



Kataloogimise automatiseerimisest Tartu Ülikooli raamatukogus;

Mis toimub Tartu Ülikooli raamatukogu hoidlates?;

Milleks veel "Mana", kui "Tulimuld" olemas? jmn.

Avaveerg	Ivi Eenmaa	4
TEADUSRAAMATUKOGU		
Et elule mitte jalgu jääda	Urve Tõnnov	5
Kataloogimise automatiseerimisest		
Tartu Ülikooli raamatukogus	Elsa Loorits, Sirje Nilbe	9
Dokumentide sisu avamine ja aineotsing arvutikataloogis	Sirje Nilbe	12
Mis toimub TÜ raamatukogu hoidlates?	Urmas Kokassaar, Kurmo Konsa	14
Eesti Riigikaitse Akadeemia raamatukogu	Sigrid Mandre	17
Eesti Kirjandusmuuseumi Arhiivraamatukogu raamatuid	Mihkel Volt	18
KULTUURISUHTED		
Leedu rahvaraamatukogude tänapäev	Vitautas Gudaitis	19
JUTUTUBA		
	Maire Liivamets	21
ARVUSTUS		
Milleks veel "Mana", kui "Tulimuld" olemas?	Maie Kalda	22
Eesti õigusbibliograafia nimestikku sirvides	Arvo Junti	24
TEIST JA TEILE		
Kukeristi Kaalu ja kirjandusuuriija	Jaanus Vaiksoo	25
Lugeja arvab:		
Kutseksam on raamatukogus töötamise garant	Mati Muru	26
JURISTIVEERG		
Ülevaade töölepingu seadusest	Heli Naeris	27
KOMMENTAAR		
Kooliajakirjadest	Hans Jürman	28
Täienduseks & täpsustuseks	Olev Esna	28
MÄLESTAME		
Kalju Oja	Hans Jürman	30
Leida Püss	Mari Sibul	31
SÕNUMID		
Asutati Eesti Hoiuraamatukogu		31
Eesti kooliraamatukogud 1994. aastal		36
TASULINE LEHEKÜLG		
Tartu Ülikooli Raamatukogus koostatavad bibliograafiakartoteegid		37
SUMMARY		
		38
PE3IOME		
		39

DOKUMENTIDE SISU AVAMINE JA AINEOTSING ARVUTIKATALOOGIS

SIRJE NILBE

Tartu Ülikooli Raamatukogu

"Kui mina ütlen mõne sõna," teatas Muna Muna, "siis tähendab see ka täpselt seda, mida mina soovin – ei rohkem ega vähem."

"Minu arust on peamine küsimus selles," lausus Alice kahtlevalt, "kas te saate sõnale nii palju erinevaid tähendusi anda."

"Minu arust on peamine küsimus," ütles Muna Muna, "kes on peremees ja kogu moos."

Lewis Carroll "Alice Peeglitagusel maal"

Sõnade tähenduse üle otsustamine moodustab suure osa tööst, mida raamatukogus teevad trükiste sisu avajad liigitajad ja märksõnastajad. Ja siis tulevad lugejad oma vajaduste ja arusaamadega, mis ei pruugi alati kokku langeda raamatukoguhoidja mõtetega. Probleem, kuidas avada dokumentide sisu nii, et lugeja aineotsing oleks tulemusrikas, on igipõline, kuid elektronandmebaaside loomine on avanud selle lahendamiseks mitmeid uusi võimalusi ja ärgitanud ümber hindama vanu.

Elektrooniliste andmebaaside ajastul tegeldakse kahtlemata rohkem sõnaliste sisuavamis- ja otsimeetoditega kui liigitusega. Siiski sõltub sisuavamismeetodite valik üleminekul kaartkataloogilt arvutikataloogile (OPACile) suuresti iga maa ja raamatukogu traditsioonidest. Arvutikataloogide loomise algetapil oli nende põhiülesanne võimaldada kindlaks teha, kas dokument on või ei ole raamatukogus olemas, seega siis bibliograafiaotsing. Tähelepanu pöörati peamiselt kirjereeglite väljatöötamisele ja kataloogide standardiseerimisele, aineotsing neis jäi tavaliselt teisejärguliseks asjaks.

Paljudes esimese põlvkonna OPACides pole aineotsing üldse võimalik või siis on otsialasid vähe ja otsing peab toimuma rangelt määratud viisil ja järjekorras. Kasutada saab ainult eelkoordineeritud otsitermineid nagu kaartkataloogis. (Eelkoordinatsioon tähendab, et liitmõisteid tähistavad märksõnad ja liigiindeksid moodustatakse indekseerimise ajal ja lisatakse kirjele tervikuna; näit. *keskaja kunst*, 159.9(075).) Abi otsiterminite ja otsistrateegia valimisel ja tulemuse kontrollimise võimalust arvuti ei paku. Väljundiks saadakse enamasti lühikirjed, mille täiendamiseks tuleb üle minna bibliograafiaotsingule.

Teise põlvkonna kataloogides on aineotsing võtme- ja märksõnade abil juba üldine. Otsitermineid saab katkestada ja Boole'i operaatorite abil ühendada. (See on **järelkoordinatsioon**, mis tähendab, et liitmõistest otsitava kohta moodustatakse alles otsingu ajal; näit. *keskaeg JA kunst*, 159.9 JA (075). Indekseerimisel on kasutatud liitmõisteid, mis otsingu käigus seotakse operaatorite JA, VÕI, EI abil. (Olulisimaid erinevusi kaart- ja elektronkataloogi vahel ongi see, et kaartkataloogis saab kasutada ainult eelkoordinatsiooni, järelkoordinatsioon on aga elektronkataloogi mägumaa.) Teise põlvkonna OPACides töödeldakse liigiindekseid nende süstemaatilist tähendust arvestades. Sobiva otsitermini leidmiseks on võimalik lehitseda registreid, küsida leitud viidete arvu ja soovi korral otsingu käigus tagasi liikuda. Dialoogis arvutiga võib valida menüü- ja käsurežiimi vahel. Arvutis on harjutusprogrammid ja eksimustest otsingu käigus antakse teada koos parandusjuhistega. Kirjeid saab väljastada mitmes formaadis. Enamik praegu töötavatest arvutikataloogidest kuulubki teise põlvkonda.

Võib rääkida ka arvutikataloogide kolmandast põlvkonnast. Need on enamasti veel katsetamisjärgus seeriatootmisel jõudmata süsteemid, milles püütakse aineotsingu võimalusi üha rohkem täiustada. Neis on võimalik esitada päringuid loomulikus keeles. Dialoog arvutiga on kasutajasõbralik, otsingu igas järgus on tagasiside hea. Sisuvamis- elementid seotakse otsingu käigus automaatselt. Kirjeid on täiendatud lisaandmetega, näit. võtmesõnadega sisukorrast ja aineloomisest. Väljundkirjed on mõnikord reastatud vastavalt nende relevantsusele.

Niisiis on arvutikataloogid aineotsingu võimaluste poolest üha lähenenud viiteandmebaasidele, kus see otsinguliik on algusest peale olnud esmatähtis. Kuid raamatukogukataloogide ja viiteandmebaaside vahele jäävad ikkagi mõned arvestamist nõudvad erinevused.

1. Viiteandmebaasides sisalduvat vahendab väljaõppinud infotöötaja. OPACi otsekasutajad on spetsiaalse koolitusega ja sageli vähese kogemusega infotarbijad - lugejad. Uuringute andmeil teevad aineotsinguid kataloogis rohkem just need lugejad, kes on vähem vilunud orienteeruma teadmiste süsteemis: nooremate kursuste üliõpilased, kooliõpilased, muidu huvilised. Teadlased, magistrandid ja õppejõud otsivad enamasti neile juba ette teada olevate teavikute kirjeid.

2. Kataloogi kirjed ei sisalda teaviku referaati, mistõttu võtmesõnade rakendamine on kataloogis piiratud ning kogude sisu avamisel tuleb pöörata tähelepanu peamiselt märksõnastamisele ja liigitusele. Võtmesõnade kasutamine ei ole tõhus ka paljumeelsete kogude kataloogimisel.

3. Viiteandmebaasid on loodud enamasti erialati, OPACid sisaldavad aga informatsiooni paljudelt aine- ja erialadelt. Seetõttu tuleb kataloogis kasutada universaalseid liigitussüsteeme ja märksõnastikke, viiteandmebaasides aga erialalisi.

4. Kataloog sisaldab valdavalt monograafiate ja jätkväljaannete kirjeid, viiteandmebaas valdavalt analüütilisi kirjeid.

Uurimused näitavad, et üleminekul kaartkataloogilt arvutikataloogile aineotsingute osa võrreldes bibliograafiliste otsingutega suureneb, moodustades vähemalt poole kõigist otsingutest. Ja kõige keerulisem on lugeja jaoks just aineotsing, mille aluseks on dokumendi kirjetele lisatud (või nendest automaatselt tuletatud) sisuelemendid - liigiindeksid, märksõnad, võtmesõnad.

Kaartkataloogis on liigitamine ja märksõnastamine alternatiivsed sisuavamismeetodid. Sisukataloogi korraldamiseks on tulnud valida kas üks või teine. Elektronkataloogis on võimalik neid kasutada nii, et nad täiendavad teineteist, võttes arvesse mõlema eripära ning tugevaid ja nõrku külgi. Ideaalis peaksid liigi- ja märksõnaotsingu võimalused ning üleminek ühe rakenduselt teisele olema võrdselt hästi läbi töötatud.

Liigituse ja märksõnastuse põhi- erinevused on järgmised:

1. Liigituse süstemaatiline lähenemine koondab materjali, võimaldab saada kogudest ülevaate erialade kaupa. Märksõnastuse alfabeetiline lähenemine hajutab materjali.

2. Liigitus on jäik. Uute teadusharude ja mõistete kajastamine liigitussüsteemides tekitab probleeme, sest nende uuendamine on aeglane protsess. Märksõnastus on paindlik. Uusi mõisteid saab süsteemi hõlpsasti lisada.

3. Liigitus võimaldab infotarbijal paremini saada ülevaadet tema jaoks uuest ainealast. Märksõnastus sobib paremini konkreetsete päringute jaoks.

4. Liigituse ekspressiivne notatsioon võimaldab päringut hõlpsalt laiendada või kitsendada. Märksõnastikus erialamõistete hierarhilised suhted ei kajastu või kajastuvad vaid tesauruse pisihierarhiates.

5. Liigitus ei sõltu keelest. Märksõnastusega kaasnevad mitmesugused keeleprobleemid: kogude paljukeelsus, terminoloogia olukord (eesti keele) automaattöötlus, infotarbija keeleoskus.

6. Liigitus opereerib pikkade abstraktsete liigiindeksitega. Märksõnastus kasutab loomuliku keele sõnavara.

7. Liigitussüsteem sobib enamasti ka teavikute paigutussüsteemiks. Märksõnad ei kõlba vähegi suuremate kogude paigutuse aluseks.

8. UDK, millega Eesti teadusraamatukogud on harjunud, saame suhteliselt odavalt valmina kätte. Märksõnastik(ud) tuleb meil endal välja arendada.

9. UDK liigitussüsteemi saab kasutada nii rahvuslikus kui rahvusvahelises andmevahetuses. Märksõnastust on isegi Eesti piires raske standardiseerida.

Liigitussüsteemid ja märksõnastikud (ka tesaurused) kuuluvad mõlemad nn. **kontrollitud** sõnavaraga infokeelte hulka. Nendele võib vastandada vaba märksõnastamise ja vabatekstiotsingu, mille hiilgeaeg, vaatamata nende meetodite odavusele ja kasutajasõbralikkusele, näib möödunud olevat, sest tänapäeva integreeritud andmebaaside kirjetel suur arv ja rahvusvaheline kasutus nõuab nende sisu täpset avamist ja ranget

kontrolli selleks rakendatava sõnavara üle. Seepärast jääb liigitamine ja märksõnastamine ka arvutikataloogide ajastul kõrget kvalifikatsiooni nõudvaks intellektuaalseks tööks. Sõnavara kontrolli vajadust alahinnates võime sattuda samasugusesse õnnetusse olukorda nagu Muna Muna:

"Kui ma ühel sõnal nii palju tähendada lasen," lisas Muna Muna, "maksan ma alati lisatasu."

"Jaa, sa peaksid nägema, kuidas nad tunglevad mu ümber laupäeva õhtuti, kui ma neile palka maksan."

KIRJANDUST

1. Chan, L. M. Subject analysis tools online: the challenge ahead // *Information Technology and Libraries*. 1990. 3. - P 258-262.

2. Gödert, W., Horny, S. The design of subject access elements in Online Public Access Catalogs // *International Classification*. 1990. 2. - P 66-76.

3. Lancaster, F. W. Vocabulary control for information retrieval. - 2nd ed. - Arlington, 1986. 270 p.

4. Larson, R. R. Between Scylla and Charybdis: Subject searching in the online catalog // *Advances in Librarianship*. Vol. 15. - San Diego etc., 1991. - P 175-236.

5. Lawrence, G. S. System features for subject access in the online catalog // *Library Resources and Technical Services*. - 1985. 1. - P 16-33.

6. Markey, K. Subject searching in library catalogs: before and after the introduction of online catalogs. - Dublin, 1984. 172 p.

7. Sisällönkuvailun organisointi Suomen kirjastoissa / P. Eskola, etc. // *Signum*. - 1991. 3. - S. 59-63.

8. Svenonius, E. Use of classification in online retrieval // *Library Resources and Technical Services*. - 1983. 1. - P 76-80.